Toxicity Based Criteria for Receiving Water Quality: Stage 2

R&D Technical Report
P312

Toxicity Based Criteria for Receiving Water Quality: Stage 2

R&D Technical Report P312

R Fleming, I Johnson (WRc), G Allen Burton (Wright State University) and M Crane (Royal Holloway, University of London)

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 165 3

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 report describes research, commissioned by the Environment Agency and SNIFFER, to test the suitability of ecotoxicological methods for application to assessing the quality of receiving water environments. The selection of methods was conducted as an earlier phase of the work (Agency R&D Note 451). The report is intended for consideration by staff in the Environmental Strategy & Environmental Protection directorates and by regional and area staff involved in environmental monitoring. The role and application of ecotoxicological methods for receiving water monitoring will be further developed, together with technical guidance, in R&D project E1D(99)03.

Research contractor
This document was produced under R&D Project 703 by:

WRc plc
Henley Road
Medmenham
Marlow
Buckinghamshire
SL7 2HD

Tel: 01491 571531 Fax: 01491 579094

WRc Report No: EA 4400/10018-1

Environment Agency Project Leader
The Environment Agency's Project Leader for R&D Project 703 was:
Dave Forrow/Oliver Warwick – National Centre for Ecotoxicology and Hazardous Substances

SNIFFER's Project Manager for R&D Project 703 was:
Peter Holmes (SEPA)

CONTENTS

Page

LIST OF TABLES ii

LIST OF FIGURES vi

EXECUTIVE SUMMARY 1

KEY WORDS 3

1. INTRODUCTION 5

1.1 Background to the study 5
1.2 Value of an integrated approach 8
1.3 Initial phases of the study 10
1.4 Current study 12

2. STUDY PROGRAMME 17

2.1 Sampling sites 17
2.2 Study design 22
2.3 Data analysis 38

3. OVERALL ASSESSMENT OF THE BIOASSAYS AND BIOMARKERS 43

3.1 Summary of the results for each method 43
3.2 Correlations between bioassays 106
3.3 Consideration of the suitability of tests for monitoring receiving water quality 111
3.4 Power of the bioassay procedures 129
3.5 Comparison of ecotoxicological data and biological survey information 134
3.6 Summary of the suitability of bioassays for receiving water monitoring 151

4. APPLICATION OF ECOTOXICOLOGICAL BIOASSAYS TO RECEIVING WATER MONITORING 157

4.1 Introduction 157
4.2 Requirements of bioassays to be used in different applications 158

5. ISSUES RELATED TO THE USE OF BIOASSAYS IN A PARTICULAR OPERATIONAL ROLE 163

Page

5.1 Introduction 163
5.2 Temporal variability 163
5.3 Small scale spatial variability 165
5.4 Power of the bioassays to detect differences in response 166

6. CONCLUSIONS 169

REFERENCES 171

APPENDICES

APPENDIX A EXAMPLE OF A CLASSIFICATION SCHEME USING BIOASSAY DATA 175

LIST OF TABLES

Table 1.1 Summary of the advantages and limitations of the chemical analytical and macroinvertebrate survey approaches 6

Table 1.2 Information provided by differential triad responses (after Chapman *et al* 1992). 8

Table 1.3 Summary of the merits and limitations of laboratory-based and *in situ* bioassays for monitoring receiving water quality 9

Table 1.4 Summary of methods recommended in the Phase 2-Stage 1 report for assessing receiving water column toxicity 11

Table 1.5 Summary of methods recommended in the Phase 2-Stage 1 report for assessing receiving water sediment toxicity 11

Table 1.6 Summary of the recent (1995) biological survey data for the study sites (ND=No data) 12

Table 1.7 Summary of toxicity data for effluents discharging at the positive control stations of the River Aire and Tees Estuary 13

Table 2.1 Location of sampling stations 17

Table 2.2 Summary information on the nature (slope, width and average depth) of the stations at the freshwater sites 22

Table 2.3 Summary of the monitoring programme 23

Table 2.4 Summary of the monitoring approach used at the River Aire site 24

Table 2.5 Summary of the monitoring approach used for the Willow Brook site 25

Table 2.6 Summary of the monitoring approach used at the Tees Estuary site 26

Table 2.7 Summary of the monitoring approach used at the Mersey Estuary site 27

Table 2.8 Summary information on laboratory test methods (For further details see Project Record P2/i701/2) 29

Page

Table 2.9 Summary of organisms and equipment needed for the laboratory-based bioassays on freshwater water column and sediment samples (for further details see Project Record P2/i701/2) 30

Table 2.10 Summary of organisms and equipment needed for the laboratory-based bioassays on estuarine water column and sediment samples (for further details see Project Record P2/i701/2) 31

Table 2.11 Summary information on *in situ* bioassays 32

Table 2.12 Summary of organisms and equipment needed for *in situ* deployments 33

Table 2.13 Transformations used in the analysis of variance of bioassay data 40

Table 3.1 Summary of the physico-chemical data for water samples taken at the different stations at each site during the study period 43

Table 3.2 Results of ANOVA for the ECL bioassay data from the River Aire and Willow Brook and the ranking of responses 45

Table 3.3 Results of ANOVA for the Microtox - acute bioassay data from the River Aire and Willow Brook and the ranking of responses 48

Table 3.4 Results of ANOVA for the *Lemna minor* growth inhibition bioassay data from the River Aire and Willow Brook and the ranking of responses 51

Table 3.5 Comparison of *Lemna minor* growth in the laboratory reference controls with the receiving water samples from the River Aire and Willow Brook 52

Table 3.6 Results of ANOVA for the *Daphnia magna* reproduction bioassay data from the River Aire and Willow Brook and the ranking of responses 54

Table 3.7 Lethality of *Daphnia magna* in River Aire receiving water samples after different exposure periods 56

Table 3.8 Summary of juvenile production in the *D. magna* reproduction tests at the different River Aire and Willow Brook stations relative to that in groundwater controls 56

Table 3.9 Results of ANOVA for the *Chironomus riparius* lethality bioassay data from the River Aire and Willow Brook and ranking of responses 58

Table 3.10 Summary of physico-chemical determinands measured in River Aire sediments on the second sampling occasion 59

Table 3.11 Summary of physico-chemical determinands measured in Willow Brook sediments on both sampling occasions 60

Table 3.12 Ranges of physico-chemical factors which support 10 day *Chironomus* survival 61

Table 3.13	Correlation coefficients (and significance) between 10 day lethality data and sediment physico-chemical parameters	62
Table 3.14	Results of ANOVA for the *Chironomus riparius* lethality development and growth data from the River Aire and the ranking of responses	63
Table 3.15	Results of ANOVA for the ECL bioassay data from the Tees and Mersey estuaries and the ranking of responses	67
Table 3.16	Results of ANOVA for the Microtox - acute bioassay data from the Tees and Mersey estuaries and the ranking of responses	68
Table 3.17	Results of ANOVA and Tukey tests for the OEL bioassay data from the Tees and Mersey estuaries and the ranking of responses and ranking of responses	72
Table 3.18	Results of ANOVA for the *Tisbe battagliai* reproduction bioassay data from the Tees and Mersey estuaries and ranking of responses	75
Table 3.19	Summary of juvenile production in the *T. battagliai* reproduction tests at the different Tees and Mersey Estuary stations relative to that in seawater controls	77
Table 3.20	Results of ANOVA for the *Corophium volutator* lethality bioassay data from the Tees and Mersey estuaries and the ranking of responses	79
Table 3.21	Summary of data from 10d *C. volutator* lethality tests carried out on sediments from the Tees Estuary	80
Table 3.22	Summary of physico-chemical determinands measured in Tees Estuary sediments on both sampling occasions	81
Table 3.23	Summary of physico-chemical determinands measured in Mersey Estuary sediments on both sampling occasions	81
Table 3.24	Ranges of physico-chemical factors which support 10 day *Corophium* survival	82
Table 3.25	Correlation coefficients (and significance) between 10 day lethality data and sediment physico-chemical parameters	83
Table 3.26	Results of ANOVA for the *Corophium volutator* growth/ lethality bioassay data from the Tees Estuary and the ranking of responses	85
Table 3.27	Results of ANOVA for the *Corophium volutator* growth/ lethality bioassay data from the Mersey Estuary and the ranking of responses	86
Table 3.28	Results of ANOVA for the *Arenicola marina* cast formation/lethality data from the Tees Estuary and the ranking of responses	88
Table 3.29	Correlation coefficients (and significance) between 10 day cast formation and lethality data in the *Arenicola marina* bioassay and sediment physico-chemical parameters	90

Page

Table 3.30 Results of ANOVA for the *G. pulex in situ* bioassay data from the River Aire and Willow Brook and ranking of responses 93

Table 3.31 Results of ANOVA for the *Daphnia magna in situ* bioassay results from the River Aire and ranking of responses 95

Table 3.32 Results of ANOVA for the *Chironomus riparius in situ* bioassay data from the River Aire and ranking of responses 96

Table 3.33 Results of ANOVA for the *Mytilus edulis in situ* bioassay data from the Tees and Mersey estuaries and the ranking of responses 98

Table 3.34 Results of ANOVA for the *Corophium volutator in situ* bioassay data from the Tees Estuary and ranking of responses 100

Table 3.35 Correlation coefficients (and significance) between responses of different laboratory based freshwater water column bioassays 107

Table 3.36 Correlation coefficients (and significance) between responses of different laboratory based estuarine water column bioassays 108

Table 3.37 Correlation coefficients (and significance) between responses of different laboratory based freshwater sediment bioassays 108

Table 3.38 Correlation coefficients (and significance) between responses of different laboratory based estuarine sediment bioassays 110

Table 3.39 Summary of the data for freshwater bioassays used at reference Station 1 and positive control Station 5 on the River Aire 112

Table 3.40 Summary of the statistical analysis of the freshwater bioassay data for Stations 1 and 5 on the River Aire 113

Table 3.41 Summary of the data for estuarine bioassays used at reference Station 8 and positive control Station 5 on the Tees Estuary 116

Table 3.42 Summary of the statistical analysis of the estuarine bioassay data for Stations 8 and 5 on the Tees Estuary 117

Table 3.43 Summary of the results of statistical analyses carried out on the bioassays used on the River Aire 120

Table 3.44 Summary of the results of statistical analyses carried out on the bioassays used on the Willow Brook 121

Table 3.45 Summary of the results of statistical analyses carried out on the bioassays used on the Tees Estuary 125

Table 3.46 Summary of the results of statistical analyses carried out on the bioassays used on the Mersey Estuary 126

Table 3.47 Summary of the suitability of freshwater bioassays for receiving water monitoring based on sensitivity and discrimination 127

Table 3.48 Summary of the suitability of estuarine bioassays for receiving water monitoring based on sensitivity and discrimination 128

Table 3.49 Summary of the Minimum Detectable Differences for freshwater bioassays and the implications for the power of the test 132

Table 3.50 Summary of the Minimum Detectable Differences for estuarine bioassays and the implications for the power of the test 133

Table 4.1 Summary of the interpretation of *in situ* and sediment bioassays (⌋ - effect, x - no effect) 160

Table 5.1 Correlation coefficients (and significance) between bioassay responses in spot and composite water column and sediment samples from the case study sites 167

LIST OF FIGURES

Figure 2.1 Sampling/deployment stations on the River Aire 18

Figure 2.2 Sampling/deployment stations on the Willow Brook 19

Figure 2.3 Sampling/deployment stations on the Tees Estuary 20

Figure 2.4 Sampling/deployment stations on the Mersey Estuary 21

Figure 3.1 Mean ECL responses on four sampling occasions for the River Aire (error bars represent one standard deviation of station samples from the mean) 44

Figure 3.2 Mean ECL responses three sampling occasions for the Willow Brook (error bars represent one standard deviation of station samples from the mean) 44

Figure 3.3 Mean Microtox - acute responses on four sampling occasions for the River Aire (error bars represent one standard deviation of station samples from the mean) 46

Figure 3.4 Mean Microtox - acute responses three sampling occasions for the Willow Brook (error bars represent one standard deviation of station samples from the mean) 47

Figure 3.5 Mean *Lemna minor* growth inhibition on two sampling occasions for the River Aire (error bars represent one standard deviation of station samples from the mean) 49

Figure 3.6 Mean *Lemna minor* growth inhibition on two sampling occasions for the Willow Brook (error bars represent one standard deviation of station samples from the mean) 50

Figure 3.7 Mean *Daphnia magna* juvenile lethality on four sampling occasions for the River Aire (* indicates no lethality measured) 52

Figure 3.8 Mean *Daphnia magna* juvenile production on four sampling occasions for the River Aire (error bars represent one standard deviation of station samples from the mean, no production occurred where there was complete lethality) 53

Figure 3.9 Mean *Daphnia magna* juvenile production on three sampling occasions for the Willow Brook (error bars represent one standard deviation of station samples from the mean) 53

Figure 3.10 Mean lethality (10 d) of *Chironomus riparius* on two sampling occasions for the River Aire (error bars represent one standard deviation of station samples from the mean, NS = no samples taken) 57

Figure 3.11 Mean lethality (10 d) of *Chironomus riparius* on two sampling occasions for the Willow Brook (error bars represent one standard deviation of station samples from the mean, NS = no samples taken) 57

Figure 3.12 Mean lethality (28 d) and development of *Chironomus riparius* on one sampling occasion for the River Aire (error bars represent one standard deviation of station samples from the mean, NS = no samples taken) 63

Figure 3.13 Mean growth (as dry weight) of *Chironomus riparius* on one sampling occasion for the River Aire (error bars represent one standard deviation of station samples from the mean, NS = no samples taken) 63

Figure 3.14 Mean ECL responses on four sampling occasions for the Tees Estuary (error bars represent one standard deviation of station samples from the mean) 65

Figure 3.15 Mean ECL responses on three sampling occasions for the Mersey Estuary (error bars represent one standard deviation of station samples from the mean) 66

Figure 3.16 Mean Microtox - acute responses on four sampling occasions for the Tees Estuary (error bars represent one standard deviation of station samples from the mean) 68

Figure 3.17 Mean Microtox - acute responses three sampling occasions for the Mersey Estuary (error bars represent one standard deviation of station samples from the mean) 69

Figure 3.18 Mean OEL responses on four sampling occasions for the Tees Estuary (error bars represent one standard deviation of station samples from the mean) 70

Figure 3.19 Mean OEL responses on three sampling occasions for the Mersey Estuary (error bars represent one standard deviation of station samples from the mean) 71

Figure 3.20 Mean *Tisbe battagliai* lethality on four sampling occasions for the Tees Estuary (* indicates no lethality measured) 73

Figure 3.21 Mean *Tisbe battagliai* juvenile production on four sampling occasions for the Tees Estuary (error bars represent one standard deviation of station samples from the mean) 73

Figure 3.22 Mean *Tisbe battagliai* lethality on three sampling occasions for the Mersey Estuary (* indicates no lethality measured) 74

Figure 3.23 Mean *Tisbe battagliai* juvenile production on three sampling occasions for the Mersey Estuary (error bars represent one standard deviation of station samples from the mean) 74

Figure 3.24 Mean lethality (10 d) of *Corophium volutator* on two sampling occasions for the Tees Estuary (error bars represent one standard deviation of station samples from the mean, NS = no samples taken) 78

Figure 3.25 Mean lethality (10 d) of *Corophium volutator* on two sampling occasions for the Mersey Estuary (error bars represent one standard deviation of station samples from the mean) 78

Figure 3.26 Mean lethality (28 d) of *Corophium volutator* for the Tees Estuary (error bars represent one standard deviation of station samples from the mean, NS = no samples taken) 83

Figure 3.27 Mean growth (as dry weight) of *Corophium volutator* for the Tees Estuary (error bars represent one standard deviation of station samples from the mean, NS = no samples taken) 84

Figure 3.28 Mean lethality (28 d) of *Corophium volutator* for the Mersey Estuary (error bars represent one standard deviation of station samples from the mean) 84

Figure 3.29 Mean growth (as dry weight) of *Corophium volutator* for the Mersey Estuary (error bars represent one standard deviation of station samples from the mean) 85

Figure 3.30 Mean lethality of *Arenicola marina* on two sampling occasions for the Tees estuary (error bars represent one standard deviation of station samples from the mean . NS = no samples taken, * indicates no lethality measured) 87

Figure 3.31 Mean cast formation of *Arenicola marina* on two sampling occasions for the Tees estuary (error bars represent one standard deviation of station samples from the mean) 88

Figure 3.32 Mean lethality of *Gammarus pulex* during two *in situ* deployments in the River Aire (ND = not deployed) 90

Figure 3.33 Mean feeding rate of *Gammarus pulex* during two *in situ* deployments in the River Aire (error bars represent one standard deviation of cage replicates from the mean, ND = not deployed) 91

Figure 3.34 Mean lethality of *Gammarus pulex* during two *in situ* deployments in the Willow Brook (ND = not deployed) 91

Figure 3.35 Mean feeding rate of *Gammarus pulex* during two *in situ* deployments in the Willow Brook (error bars represent one standard deviation of cage replicates from the mean, ND = not deployed) 92

Figure 3.36 Mean lethality of *Daphnia magna* during two *in situ* deployments in the River Aire (error bars represent one standard deviation of cage replicates from the mean, ND = not deployed, * = no lethality measured, L = lost) 94

Figure 3.37 Mean *in situ* lethality of *Chironomus riparius* during two *in situ* deployments in the River Aire (error bars represent one standard deviation of cage replicates from the mean, ND - not deployed) 95

Figure 3.38 Mean clearance rate of *Mytilus edulis* during two *in situ* deployments in the Tees Estuary (error bars represent one standard deviation from the mean) 97

Figure 3.39 Mean clearance rate of *Mytilus edulis* during two *in situ* deployments in the Mersey Estuary (error bars represent one standard deviation from the mean) 98

Figure 3.40 Mean lethality of *Corophium volutator* during two *in situ* deployments in the Tees estuary (error bars represent one standard deviation of cage replicates from the mean, ND = not deployed, L = lost). 99

Figure 3.41 Mean levels of MT (and 95% confidence levels) of mussels from seven sites along the Mersey Estuary (Station 8 missing) 103

Figure 3.42 Mean levels of MT (and 95% confidence levels) of mussels from seven sites along the Tees Estuary (Station 5 missing) 103

Figure 3.43 Mean neutral red retention times (and 95% confidence levels) of mussels from eight sites along the Mersey Estuary 104

Figure 3.44 Mean neutral red retention times (and 95% confidence levels) of mussels from seven sites along the Tees Estuary (Station 3 missing) 104

Figure 3.45 Box and Whisper plots for median heart rate of mussels from eight sites along the Mersey Estuary 105

Figure 3.46 Box and Whisper plots for median heart rate of mussels from seven sites along the Tees Estuary (Station 3 missing) 105

Figure 3.47 Two-dimensional MDS ordination of macroinvertebrate abundance data for stations on the River Aire (numbers indicate stations) 135

Figure 3.48 Primary and derived univariate biological measures of community structure at selected stations on the Tees Estuary (means ± SD, A = abundance, T = total number of taxa, B = biomass as mg, A/T and B/A = ratios of A:T and B:A respectively) 139

Figure 3.49 Univariate measures of biological diversity at selected stations on the Tees Estuary (means ± SD, H' = Shannon-Wiener diversity index, J' = Pielou's index of eveness, 1/D = Simpson's diversity index) 140

Page

Figure 3.50 Two-dimensional MDS ordination of macroinvertebrate abundance (A) and biomass (B) data for stations on the Tees Estuary (number indicates station) 143

EXECUTIVE SUMMARY

The Environment Agency (as the NRA) identified a need for General Quality Assessment (GQA) schemes to quantify geographic differences and to monitor temporal trends in the water quality of rivers, canals, estuaries and coastal waters in England and Wales. As part of this initiative a three-stage programme was devised by the Environment Agency and SNIFFER to investigate the development of a national toxicity-based scheme for classifying water column and sediment quality in both fresh and saline water. This classification would potentially complement existing and proposed schemes based on biology, chemistry, nutrients and aesthetics. Stage 1 of the programme recommended a field study programme in which a number of laboratory and *in situ* toxicity bioassays could be applied at a number of riverine and estuarine sites in the UK in order to determine the most suitable candidates for use in a national monitoring scheme. This report covers Stage 2 of the programme in which these bioassay methods were piloted at two riverine and two estuarine sites and the results of each were assessed in terms of their sensitivity, discrimination, cost-effectiveness, practicality and ecological significance. The potential for use of these tests in local impact assessments was also considered.

Four UK locations were chosen for the Stage 2 study: the River Aire and Willow Brook freshwater systems and the Mersey and Tees estuaries. These sites were chosen because they contained stations ranging from good to bad quality based on biological and chemical monitoring data and eight sampling stations representing this graduation were identified at each site. Testing was carried out in a Summer window (July to October 1996) to represent perceived worst case conditions in terms of flow conditions and each site was sampled a minimum of three times in order to assess temporal variability. Water and sediment samples were removed from each site for laboratory-based bioassays and a variety of test organisms were also deployed *in situ*. In order to provide complementary information, researchers at Plymouth University performed a suite of biomarker tests on mussels deployed *in situ* in the two estuaries.

The testing of field collected water and sediment samples with laboratory-based water column and sediment bioassays and the deployment of *in situ* bioassays at two riverine (River Aire and Willow Brook) and two estuarine (Tees and Mersey estuaries) sites in a summer window (July to October 96) resulted in the following conclusions being drawn:

- on the basis of the sensitivity, discrimination between stations and statistical power of the bioassays a number of methods of each type can be used (in there present format or with minor modifications) to assess receiving water quality:

Type of bioassay	Suitable candidates for different environments	
	Freshwater	Estuarine
Laboratory-based water column	10 d *D.magna* reproduction	9 d *T.battagliai* reproduction or OEL
Laboratory-based sediment	10 d *C.riparius* lethality	10 d *C.volutator* lethality or 10 d *A.marina* cast formation
In situ	6 d *G.pulex* feeding rate	None currently appropriate

The bioassays given in the table above all showed a range of responses from limited effects at reference stations to marked effects at impacted stations. They were able to discriminate statistically between reference and impacted stations since in many instances the bioassays were able to separate stations on the basis of differences in response of 25% effect or lower. A number of the bioassays had both sub-lethal and lethal endpoints which enabled stations to be separated on two levels:

- comparisons of the bioassay data and macroinvertebrate survey data on the River Aire and Tees Estuary have shown that certain of the laboratory-based and *in situ* bioassays of receiving water column and sediment toxicity are predictive of patterns of macroinvertebrate assemblage structure at these highly impacted sites. Although a large number of bioassays was used in both the freshwater and estuarine studies, a smaller subset of assays was effective in predicting natural macroinvertebrate patterns of distribution and abundance. These findings have important implications for estuarine locations where there is no standardised procedure for carrying out macroinvertebrate surveys and the use of an appropriate battery of bioassays could represent a cost-effective means of monitoring receiving water quality. However, certain bioassays were also able to detect the effects of intermittent discharges or sub-lethal effects at reference locations which had not resulted in impacts on the macroinvertebrate community;

- a number of longer-term laboratory-based sediment bioassays (28 d *C.riparius* growth and 28 d *C.volutator* growth) or *in-situ* bioassays (6 d *D.magna* lethality and 10 d *C.volutator* lethality) could be potentially useful for receiving water monitoring but further standardisation is needed before they can be used routinely.

- the type of bioassays used for a particular operational role depends on the objectives of the study. For example in the case of general quality assessment where monitoring has to consider the effects of intermittent and continuous point and non-point sources the use of laboratory-based sediment bioassays and *in situ* bioassays are considered to be the most appropriate. For local environmental impact assessments of the impact of specific discharges laboratory-based water column bioassays are considered to be most appropriate.

- the use of composite water column or sediment samples represents a means of reducing the costs associated with general quality assessments whilst still accounting for small scale spatial effects at a locations.

KEY WORDS

Bioassays, *in situ*, water column, sediment, River Aire, Willow Brook, Tees Estuary, Mersey Estuary, enhanced chemiluminescence (ECL) test, Microtox, *Lemna minor, Daphnia magna, Gammarus pulex, Chironomus riparius,* oyster embryo larval development (OEL) test, *Tisbe battagliai, Corophium volutator, Arenicola marina, Mytilus edulis.*

1. INTRODUCTION

1.1 Background to the study

In the United Kingdom a combined chemical analysis and biological survey approach is currently used to assess freshwater receiving water quality and provide the information on which water quality management decisions are made. This 'comprehensive' picture of the quality of rivers and canals is provided by reporting a grading of chemical quality (based on the results of analyses of dissolved oxygen, ammonia, biochemical oxygen demand and certain trace metals) and biological quality (based on the monitoring of benthic macroinvertebrates). Each approach has its advantages and limitations and these are summarised in Table 1.1 using information given in Hellawell (1986), Metcalfe-Smith (1994)

Recently the Environment Agency (as the NRA) identified a need for General Quality Assessment (GQA) schemes to quantify geographic differences and to monitor temporal trends in the water quality of rivers, canals, estuaries and coastal waters in England and Wales. GQA will enable water quality management and control to be cost-effectively focused and the consequences of such actions to be assessed. Such schemes for basic chemical quality and for biological quality already exist for use in canals and rivers as described above. The Environment Agency's Development and Testing of General Quality Assessment Schemes Project (No: 469) extended the scope of the rivers GQA to include nutrients and aesthetics and developed a new scheme for estuaries and coastal waters based on basic chemistry, nutrients, aesthetics and sediment quality. Similar schemes have also been developed by ADRIS for application in Scotland and Northern Ireland. However, UK regulators are also interested in having classification schemes for receiving waters based on biological effects testing (direct toxicity assessment) to complement the chemical analysis and biological survey based schemes. The direct toxicity assessment approach is particularly relevant for estuarine and coastal waters where there are no standardised biological survey procedure currently in place. A potential role for receiving water biological effects measures in the local environmental impact assessments of complex effluents discharged to fresh or marine waters has also been identified as part of the application of toxicity-based limits (Wharfe 1996).

However, as well as being a 'quality measure' biological effects testing also has a role in investigating cause and effect relationships which link together analytical chemistry data with that from biological surveys carried out in GQA schemes and local environmental impact assessments. Biological effects measures are also useful in identifying sources of pollution and in differentiating between water column and sediment toxicity so it can be determined whether changes in point source control have resulted in improvements in water quality and/or whether problems with contaminated sediments remain to be addressed.

Table 1.1 Summary of the advantages and limitations of the chemical analytical and macroinvertebrate survey approaches

Approach	Merits	Limitations
Chemical analysis	Test methods are well established and can be carried out to standard operating procedures thereby generating repeatable and reproducible data Testing is cost-effective for a limited suite of well-defined determinands	The determinands measured are only a small subset of potential contaminants and will not consider breakdown products. Ecotoxicological data are not available for many substances meaning analytical chemistry data cannot be interpreted. Biological effects may occur at concentrations below analytical limits of detection. It is not possible to establish directly the extent of any biological effect from the analytical chemistry data due to issues of bioavailability and the interaction of contaminants.
Macroinvertebrate surveys	Macroinvertebrate communities are differentially sensitive to pollutants, react to them quickly and are, therefore, capable of providing a graded response. Macroinvertebrate communities are very heterogeneous, with numerous phyla and trophic level represented and there is a strong possibility that at least some organisms will respond. Macroinvertebrates are present in most aquatic habitats, are abundant and relatively easy and inexpensive to collect. Benthic invertebrates are relatively sedentary and are representative of local conditions. Macroinvertebrates have lifespans long enough to provide a record of environmental quality.	Changes in macroinvertebrate structure indicate an impact has already occurred and do not allow preventative action. The current scheme used for UK freshwaters is based on the sensitivity of representative species to organic pollution rather than chemical contaminants. Macroinvertebrate surveys cannot be used in highly impacted sites to detect change. Macroinvertebrate surveys cannot be used in certain areas, for example, channelized sections of rivers. Macroinvertebrates respond to seemingly minor changes in physico-chemical parameters which can confound data interpretation if not taken into account. Macroinvertebrates have complex life histories and the responses of surveys can vary with season.

Approach	Merits	Limitations
	The responses of macroinvertebrate communities incorporate inter-species effects such as predator-prey relationships.	Macroinvertebrates have slow response times in terms of both impact and recovery. Spatial heterogeneity in macroinvertebrate communities is high requiring considerable replication to detect subtle changes. Changes in macroinvertebrate structure (abundance and diversity) may not result in changes in ecosystem function. There is no standardised macroinvertebrate survey scheme for estuarine/marine environments.

1.2 Value of an integrated approach

Regulators have recognised that for water quality monitoring and management biological effects data obtained from toxicity tests or bioassays can provide useful information as part of either a tiered or an integrated approach with chemistry and biology. Multivariate statistical techniques now mean that analysing complex multitype datasets has become relatively straightforward which aids in directing subsequent regulatory actions. This integrated approach has been applied to sediments in the 'quality triad' (Chapman 1990). In aquatic environments there will clearly be scenarios where the responses of the biological and ecotoxicological data might be expected to give similar responses, for example in very contaminated conditions (Birge *et al* 1989, Eagleson *et al* 1990, Dickson *et al* 1992). There will also be scenarios when the response of the biological and ecotoxicological data might be expected to be different, for example in conditions of moderate to low contamination (Dickson *et al* 1992). In these situations, the extra sensitivity of the toxicological response might be expected to give a signal but the effects might be such that they have not fed through to the community level. Therefore, regulatory action can be implemented before a signal in the biological data might have become apparent. Across the spectrum of contamination, the biological and ecotoxicological responses could provide complementary information which together provides better assessments and an enhanced view of environmental quality. Table 1.2 shows the information that can be obtained from a quality triad approach and the type of conclusions that can be drawn.

Table 1.2 Information provided by differential triad responses (after Chapman *et al* 1992).

Chemistry (Contamination)	Toxicity	Biology (Alteration)	Possible conclusions
+	+	+	Strong evidence for chemical-induced degradation
-	-	-	Strong evidence that there is no pollution-induced degradation
+	-	-	Chemical contamination is observed but the substances are not bioavailable to organisms or are present at non-toxic levels
-	+	-	Contamination with unidentified chemicals or conditions exist with the potential for degradation
-	-	+	The system is stressed by factors other than the levels of chemical contaminants
+	+	-	Toxic chemicals are stressing the system but at present are not sufficient to significantly modify the system
-	+	+	Unmeasured toxic chemicals are causing degradation which may relate to the type of sampling approach used
+	-	+	Chemicals contamination is observed but the substances are not bioavailable or degradation is not due to toxic chemicals

Responses are shown as either positive (+) or negative (-), indicating whether or not measurable (statistically significant) differences from reference conditions/measures are determined. Contamination relates to the presence of toxicants in the receiving water (water column and sediment) while degradation relates to a reduction in the abundance and diversity of macro-invertebrates

Direct biological-effect testing, where some measure of biological 'health' is determined, can aid decision-makers in these tasks by predicting and detecting effects, screening and prioritising problems and by quantitatively estimating hazards and will ultimately improve the quality and confidence of these decisions. Effect testing can, therefore, be used as a potential environmental quality 'viewpoint' and be used to set and enforce standards and measure the success of regulatory measures. Laboratory and *in situ* bioassays can provide a relatively inexpensive and rapid means of estimating environmental quality. They can be used wherever and whenever an environmental sample is collected, providing a degree of spatial and temporal definition and comparability not attainable in any other way making them useful for monitoring large areas, identifying 'hotspots' and identifying the source and cause of any impact. In areas where it may be difficult to assess biological community parameters (e.g. estuarine/marine environments) it represents a useful indicator of effect and when used together with biological community information and chemistry it helps link cause with effect and adds weight of evidence to any decision making.

The assessment of freshwater and estuarine surface water and sediment toxicity could be carried out using two approaches: laboratory-based bioassays on field collected samples and *in situ* methods. Each of the approaches has their own merits and limitations (see Table 1.3) and the use of both types of methods should be considered in any monitoring programme. *In situ* methods provide a link between controlled laboratory-based bioassays and effects on benthic assemblages in the field. They are more representative than laboratory bioassays of field conditions, since they integrate fluctuations in exposure conditions over time and they are not confounded by the problems of testing spot environmental samples which may not be representative of the receiving water.

Table 1.3 Summary of the merits and limitations of laboratory-based and *in situ* bioassays for monitoring receiving water quality

Type of test	Merits	Limitations
Laboratory-based bioassays	Control over extraneous test variables other than contaminant levels. Influence of test variability can be minimised.	Sample collection costs may be high. Testing may indicate sampling induced artefacts. Sample stability may prevent the use of sub-lethal tests which are required to detect the presence of contaminants causing mainly chronic effects.
In situ bioassays	Provide an integrated assessment of biological effects over extended periods of time. The extended nature of exposure allows the presence of contaminants causing mainly chronic effects to be detected. Responses measured tend to be of greater ecological relevance.	The exposure history causing effects is not known. Deployment systems are vulnerable to damage. The precision of the test methods may be lower than laboratory-based tests. Difficulty of not having a reference station as a control at a site.

To progress the development of an integrated approach a three-stage programme to investigate the suitability of different methods for general quality assessment (and local environmental impact assessment) was initiated by the Environment Agency and Scotland and Northern Ireland Forum For Environmental Research (SNIFFER). The scheme, once developed could potentially be used for local environmental impact assessments as well as in a national GQA scheme.

1.3 Initial phases of the study

In Phase 1 of the project a scoping study was carried out (by Environment and Resource Technology) which reviewed potential assessment methods for estuarine and coastal surface waters and sediments, taking into account the availability, practicality and relevance of methods (Environment Agency 1996a). The review covered approaches using whole organisms, biotechnology (biosensors and similar methods) and biomarkers, with the objective of formulating recommendations for the application of appropriate methods to general quality assessments and local environmental impact assessments. The review recommended that for the surface waters an initial programme of field evaluation and validation should include laboratory-based methods (using marine copepod *Tisbe battagliai* lethality, oyster embryo-larval development, bacterial bioluminescence and enhanced chemiluminescence methods) on discrete samples and *in situ* methods (using caged marine mussels). For sediments it was recommended that both the *Corophium volutator* and *Arenicola marina* sediment bioassays should be used.

In Phase 2 (Stage 1) of the programme, built on the recommendations of the Phase 1 report, a field study programme was developed in which a number of laboratory and *in situ* toxicity bioassays could be applied at a number of riverine and estuarine sites in the UK. The study identified the most suitable methods for use in a national monitoring scheme and local environmental impact assessments (Environment Agency 1997a). A combination of laboratory and *in situ* methods was recommended as each approach has its advantages and limitations (see Table 1.2). By using both types of test it would be possible to draw more meaningful conclusions as to which type of procedures were most appropriate for different applications. The data from these bioassays could be used to ascertain whether they show similar or different levels of response at a location due to the different exposure conditions and aid in identifying possible sources of contamination (that is episodic or continuous point sources or non point sources). It was considered that the most appropriate comparison of laboratory-based and *in situ* methods, would use the same species in each type of bioassay. The selection of methods considered whether tests using a particular species and endpoint could be applied to both field collected water column or sediment samples and also deployed *in situ*. This approach has been adopted by workers in the United States for assessing the toxicity of effluent and storm water discharges (Sasson-Brickson and Burton 1991, Ireland *et al* 1996), since it provides a 'common toxicological currency'. This was considered to be particularly important in a local environmental impact assessment when evaluating whether toxicity-based limits imposed on effluent discharges are achieving the desired goals (for example, no acute lethal or chronic sub-lethal toxicity) at designated points in the receiving water.

The tests shown in Table 1.4 were recommended for use in the assessment of sub-surface water toxicity and followed the recommendations of the Phase 1 report (Environment Agency 1996a).

Table 1.4 Summary of methods recommended in the Phase 2-Stage 1 report for assessing receiving water column toxicity

Type of test	Candidate test systems for each type of receiving water	
	Freshwater	Estuarine
Laboratory-based tests	Enhanced chemiluminescence Microtox acute 10 d *Daphnia magna* reproduction and lethality	Enhanced chemiluminescence Microtox acute 9 d *Tisbe battagliai* reproduction and lethality Oyster embryo-larval development test
In situ tests	6 d *Gammarus pulex* feeding rate or 6 d *Daphnia magna* lethality	28 d *Mytilus edulis* clearance rate

The tests shown in Table 1.5 were recommended for use in the assessment of sediment toxicity.

Table 1.5 Summary of methods recommended in the Phase 2-Stage 1 report for assessing receiving water sediment toxicity

Type of test	Candidate test systems for each type of receiving water	
	Freshwater	Estuarine
Laboratory-based tests	10 d *Chironomus riparius* lethality 28 d *Chironomus riparius* growth and lethality	10 d *Corophium volutator* lethality 28 d *Corophium volutator* growth and lethality 10 d *Arenicola marina* lethality and cast formation
In situ tests	10 d *Chironomus riparius* lethality	10 d *Corophium volutator* lethality

A number of the tests (*D. magna* and *T. battagliai* reproduction tests, *C. riparius* and *C. volutator* growth tests, *G. pulex in situ* test and *A. marina* cast formation test) allow multiple (lethal and sub-lethal) endpoints to be measured which in principle means that the methods can discriminate between locations at lethal and sub-lethal levels. For example certain stations at a site may be separated from the others using a lethality endpoint while among those showing limited mortality a gradient of effect may be observed for the sub-lethal endpoint.

1.4 Current study

1.4.1 Background

This report covers Phase 2 (Stage 2) of the programme in which these bioassay methods were piloted at freshwater and estuarine sites. Four UK sites were chosen for the Phase 2 (Stage 2) study: the River Aire and Willow Brook freshwater sites and the Mersey and Tees estuarine sites. These sites were chosen because they encompassed stations ranging from good to bad quality, based on both historical biological and chemical monitoring data from previous years. Eight sampling stations representing this gradient were identified at each site (see Table 1.6). On the River Aire there was a measured decline in biological quality from no pollution at location 1 (BMWP=167, ASPT=5.76) through slight pollution at location 2 (BMWP=72, ASPT=4.00), greater pollution at location 4 (BMWP=41, ASPT=3.42) and gross pollution at locations 7 and 8 (BMWP=12-20, ASPT=2.40-2.86). In terms of chemical monitoring data, location 1 was chemical class B, location 2 chemical class C while locations 4 and 7 were of class D and E respectively. On the Tees Estuary the highest number of taxa recorded in a 1995 survey were recorded at locations 1, 3 and 8 ($\geq$16 taxa), while those measured at locations 4 to 7 were lower (<9 in all cases). The pattern of biological quality for the Willow Brook is complicated by the branched structure of the Brook but was lowest at Station 1 (BMWP=6) and highest at Station 8 (BMWP=95). No recent data were available for the Mersey Estuary.

Table 1.6 Summary of the recent (1995) biological survey data for the study sites (ND=No data)

Site	Measure of biological quality	Location 1	2	3	4	5	6	7	8
River Aire	BMWP score	167	72	ND	41	ND	26	12	20
	ASPT score	5.8	4.0	ND	3.4	ND	2.9	2.4	2.9
Willow Brook	BMWP score	6	ND	ND	53	26	45	28	95
Tees Estuary	Total no. of taxa	18	ND	16	8	6	8	9	22
Mersey Estuary		ND	ND	ND	ND	ND	ND	ND	ND

In the study programme the River Aire and Tees Estuary sites included stations regarded as an unimpacted ('clean') reference location (based on chemical and biological data) and an impacted positive control location. These two stations represented opposite ends of a pollution gradient. The reference locations experienced no or limited inputs of chemical contaminants. In contrast, the impacted positive control stations were subject to high concentrations of chemical contaminants from effluent discharges which had previously been found to have the potential to impact the waters in the vicinity of the discharges based on their toxicity and the available dilution (Environment Agency 1996b). On the River Aire the unimpacted reference location was Station 1 (Bell Busk Road) while the impacted positive control location was Station 5 (in the discharge plume from Esholt STW). On the Tees Estuary Station 8 (Seal Sands) was the reference location whilst Station 5 (at the confluence of Dabholm Gut receiving ICI Wilton and Bran Sands STW discharges) was the positive control location.

Table 1.7 summarises the toxicity data for the effluents discharging to the River Aire and Tees Estuary at the positive control stations and indicates the extent of the potential receiving water impact. The data were obtained during the Environment Agency/SNIFFER funded Toxicity-Based Consents Pilot Study (Project No: 493) which investigated the application of a draft protocol for controlling complex and toxic discharges to industrial and sewage treatment works releases to fresh, estuarine and marine waters (Environment Agency 1996b). At the River Aire the Predicted No Effect Concentration (PNEC) for the Esholt sewage treatment works final effluent derived from a series of tests (72h R*aphiodocelis subcapitata* growth inhibition, 48h *Daphnia magna* immobilisation and 96h juvenile *Onchorynchus mykiss* lethality) was 4.5 Toxic Units, representing 22% v/v effluent. The limited dilution in the river at the point of discharge meant that the Predicted Environmental Concentration (PEC) of 3.3 TU was less than the PNEC indicating a risk of receiving water impact that was confirmed by biological surveys (see Table 1.7). The discharge at the positive control station on the Tees Estuary was highly toxic and the PNEC of 455 TU, representing 0.22% v/v effluent, (72h *Phaeodactylum tricornutum* growth inhibition, 24h oyster embryo-larval development and 96h juvenile *Scophthalmus maximus* lethality) was considerably higher than the PEC of 100 TU again indicating a risk of receiving water impact that was confirmed by biological surveys (see Table 1.7).

Table 1.7 Summary of toxicity data for effluents discharging at the positive control stations of the River Aire and Tees Estuary

Location	Toxicity of input at positive control station		Predicted Environmental
	Most sensitive test	PNEC (TU)	Concentration (TU)
River Aire	48h *D.magna* immobilisation	4.5	3.3
Tees Estuary	24h oyster embryo-larval development	455	100

1.4.2 Aims of the current study

The objectives of the present study were to:

- assess the ability of a series of laboratory-based and *in situ* bioassays to discriminate between sample locations at the two freshwater (River Aire and Willow Brook) and the two estuarine (Tees and Mersey estuaries) sites that according to previous biological survey data show a gradient of biological quality (see Table 1.6). In the comparison of methods other key factors that have been considered are sensitivity, cost-effectiveness, ease of use and robustness;

- assess whether the results of tests using composite samples adequately reflected those obtained with a series of discrete samples;

- determine by synoptic ecotoxicological testing and macroinvertebrate surveys whether the two approaches yield complementary data or result in the same conclusions being drawn.

The results obtained in this study have been evaluated in the context of other integrated studies using chemical, biological effects and biological survey data (for example Birge *et al* 1989, Swartz *et al* 1994, Chapman *et al* 1996) which have emphasised the importance of all the elements of the 'holistic' approach.

1.4.3 Overview of the approach

Testing was carried out in a Summer window (July to October 1996) to represent worst case conditions in terms of flow and each site was visited a minimum of three times in order to assess temporal variability. Water and sediment samples were removed from each site for laboratory-based bioassays and a variety of test organisms were also deployed *in situ*. Spatial variability was assessed by sample replication at each station. In addition to the toxicity bioassays, biological survey data and physico-chemical sediment data (total organic carbon and particle size distribution) were generated for each station in order to aid interpretation of the results and to assess the ecological significance of the test data. The underlying assumption of the study design was that biological quality, as measured by commonly used benthic indices, is an accurate reflection of 'in-stream' conditions. However, it has to be remembered that the biological survey technique currently used for freshwater systems was originally developed to assess the effects of organic loading and the extent to which it reflects gradients of chemical contamination has not been fully established.

Although the Phase 2-Stage 1 report did not recommend the use of biomarkers, an opportunity to trial some of these methods during the project arose and these were included in the study. A biomarker can be defined as "a biochemical, cellular, physiological or behavioural variation that can be measured in tissue or body fluid samples or at the level of whole organisms that provide evidence of exposure to and/or effects of, one or more chemical pollutants (and/or radiations)" (Depledge 1993). These methods represent a new approach in pollutant monitoring and have been developed to be diagnostic of toxic responses and provide an early indication of exposure/responses (Giddings 1986, Depledge 1992).

Biochemical and cellular biomarkers can be advantageous because they enable early, sub-lethal responses of organisms to be related to specific types of pollutant exposure. Physiological biomarkers integrate toxicity in different target tissues and are useful because responses can be related to one, or more, Darwinian fitness parameters, for example: growth, scope for growth, reproductive viability (Depledge and Lundebye 1996). The use of a molecular or cellular biomarker approach is limited because pathological damage to specific tissues or organs cannot always be proven. Physiological responses to chemical exposure have been largely ignored because they are regarded as too generalised. However, the combination of these techniques as a suite of biomarker tests removes many of the problems caused by their individual use.

In this study the combination of available biomarkers that were applied at the estuarine sites included analyses of metallothioneins, lysosomal integrity and cardiac activity in mussels. Metallothioneins (MT) are low molecular weight proteins capable of binding trace metals in the cells of a wide range of organisms, and represent a potential molecular biomarker. A number of trace metals (including Cu, Cd, Zn, Hg, Pt, Au, Ag and Bi) induce metallothioneins in invertebrates. It has been established that metallothioneins are involved in the binding and detoxification of non-essential trace metals as well as playing a role in the regulation of

essential metals such as copper and zinc. Field studies have shown that MT concentrations may be a useful indicator of exposure to trace metals in the aquatic environment (Hogstrand and Haux 1990, Roch *et al* 1982).

Biomarkers of cellular structure and function, induced by pollutant exposure, are considered as early warning markers of continuing, and potentially increasing pathological damage. Lysosomal instability as a consequence of chemical exposure has been extensively studied as a cellular biomarker in fish hepatocytes (Moore 1985, Kohler 1989), molluscan digestive cells (Lowe and Pipe 1996) and haemocytes (Moore *et al* 1996).

Several physiological responses of the blue mussel, *Mytilus edulis*, have been tested under laboratory conditions for their suitability to detect environmental stressors (Akberali and Trueman 1985). Those that offer potential as physiological biomarkers include respiration and filtration rates, pumping rates, valve activity and heart beat (Baldwin and Kramer 1994). Heart-rate data are useful in that they can provide an indication of the general physiological state of *Mytilus edulis*.

In the field study it was decided to attempt to measure metallothionein levels, lysosome integrity and cardiac activity in *Mytilus edulis* deployed along the estuarine sites to determine whether these biomarkers used together might be utilised with other biological effect measures in the determination of general environmental quality. The biomarker work was carried out by the University of Plymouth.

1.4.4 Reporting

This final report has been updated based on the comments received on a Draft R&D Technical Report P118/i703 produced in July 1997. It reflects recent changes in the position of regulators in the United Kingdom with respect to the use of biological effects testing for receiving water monitoring.

2. STUDY PROGRAMME

This section summarises the work programme carried out at the four case study sites. Further details are given in Project Record P2/i701/2 along with the protocols for the test methods.

2.1 Sampling sites

Table 2.1 lists the locations and the grid references of the sampling/deployment stations at each of the case study locations. Figures 2.1-2.4 show the sampling/deployment stations.

Table 2.1 Location of sampling stations

Site	Sampling stations	Grid Reference	Station Code
River Aire	Bell Busk Road	SD 903 558	Aire-1
	U/S Cononley Beck	SD 995 467	Aire-2
	D/S Marley STW	SE 093 408	Aire-3
	Crossflats	SE 095 404	Aire-4
	D/S Esholt STW*	SE 190 388	Aire-5
	Fish pond Lock	SE 357 300	Aire-6
	Castleford	SE 432 263	Aire-7
	D/S Ferrybridge	SE 483 246	Aire-8
Willow Brook	Water Lane Weldon	SP 921 896	WBrook-1
	Weldon Lodge	SP 917 915	WBrook-2
	Golf Course	SP 936 920	WBrook-3
	A6116	SP 900 887	WBrook-4
	A427	SP 930 895	WBrook-5
	Lodge Farm	SP 946 911	WBrook-6
	Blatherwycke Lake Outlet	SP 981 967	WBrook-7
	Apethorpe	TL 026 959	WBrook-8
Tees Estuary	Teesport	NZ 527 222	Tees-1
	Opposite Tees Dock	NZ 542 252	Tees-2
	U/S Dabholm Gut	NZ 549 255	Tees-3
	Opposite Dabholm Gut	NZ 545 267	Tees-4
	Dabholm Gut Confluence*	NZ 549 258	Tees-5
	D/S Dabholm Gut	NZ 549 251	Tees-6
	Bran Sands Channel	NZ 548 265	Tees-7
	Seal Sands	NZ 533 264	Tees-8
Mersey Estuary	Perch Rock	SJ 312 942	Mersey-1
	Wallasey Landing Stage	SJ 327 897	Mersey-2
	Tranmere Oil Jetty	SJ 337 869	Mersey-3
	Eastham Locks	SJ 370 811	Mersey-4
	Pier Head	SJ 336 905	Mersey-5
	Dingle	SJ 359 868	Mersey-6
	Garston	SJ 398 835	Mersey-7
	Hale	SJ 483 816	Mersey-8

* These sampling stations are located within the mixing zones of effluent discharges and are regarded as positive controls for the water column and *in situ* bioassays particularly (see Table 1.7)

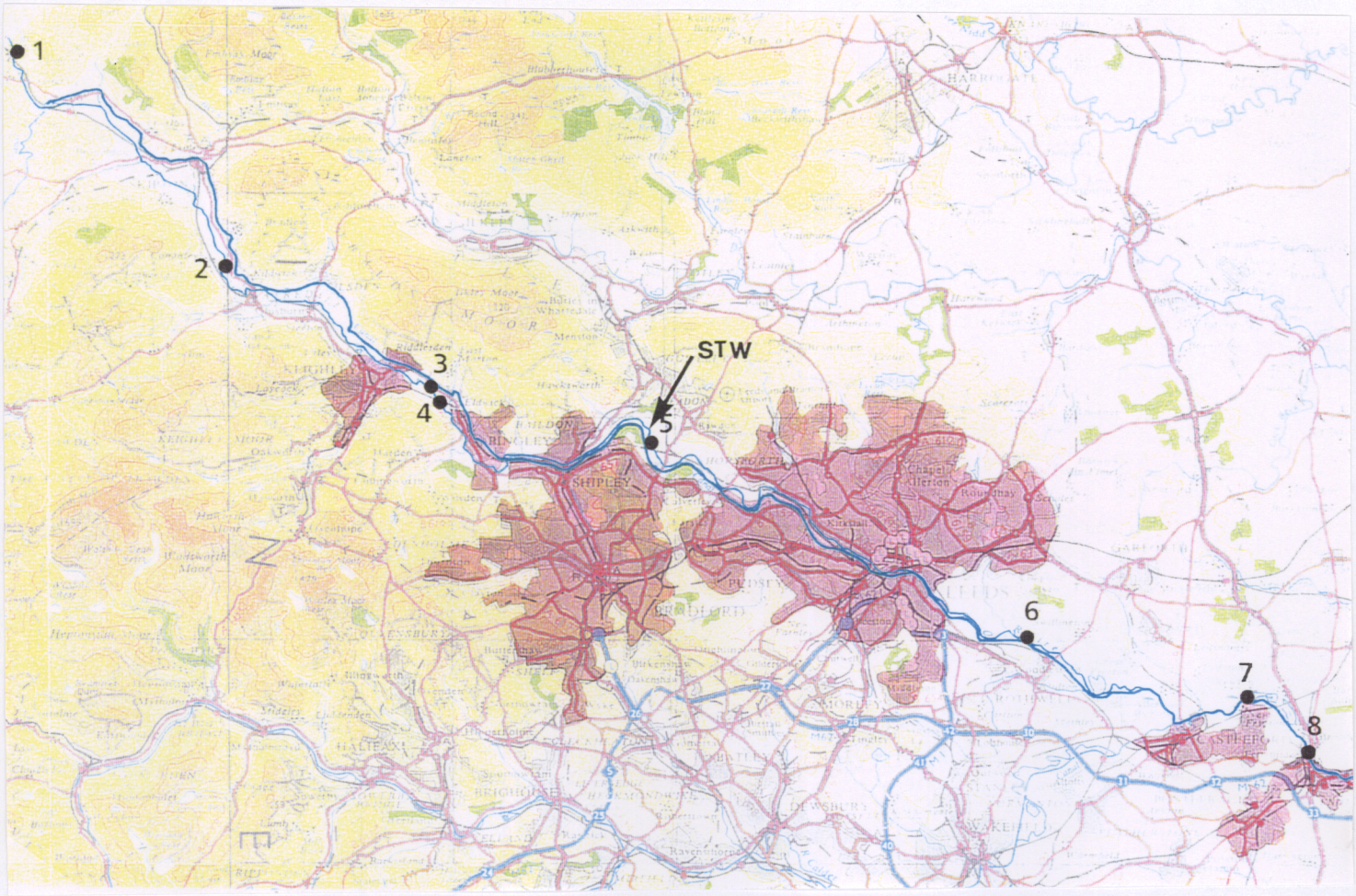

Figure 2.1 Sampling/deployment stations on the River Aire

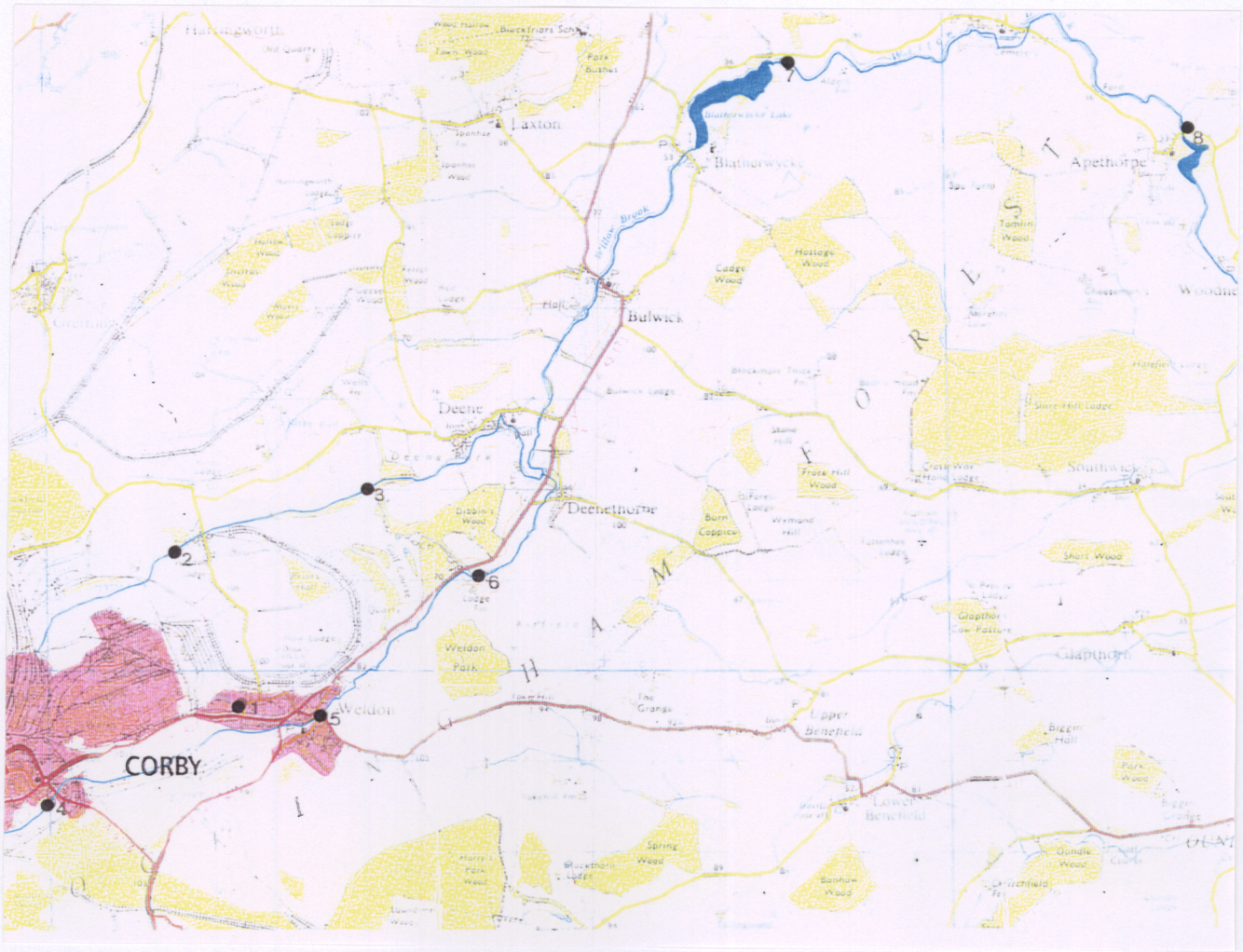

Figure 2.2 Sampling/deployment stations on the Willow Brook

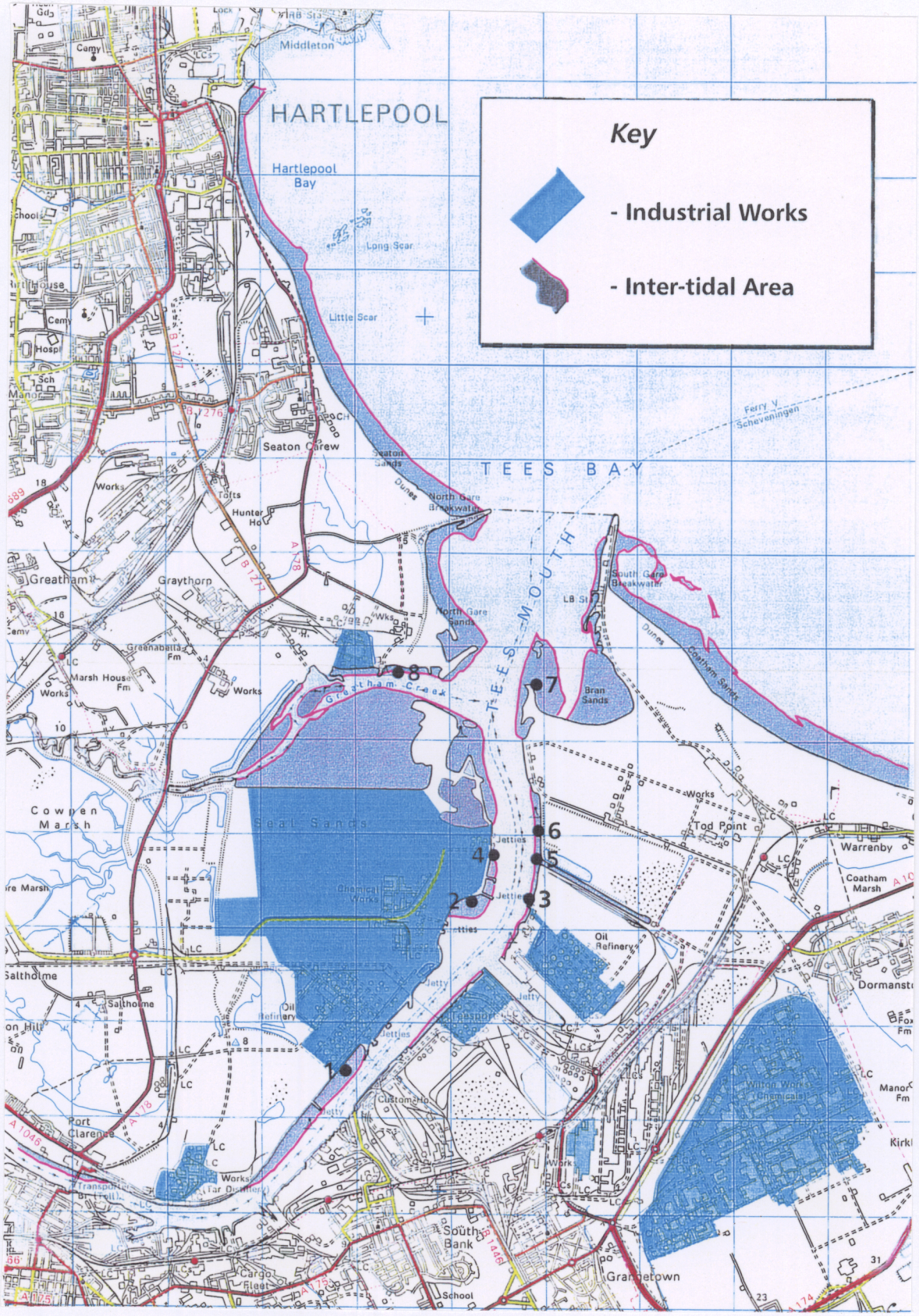

Figure 2.3 Sampling/deployment stations on the Tees Estuary

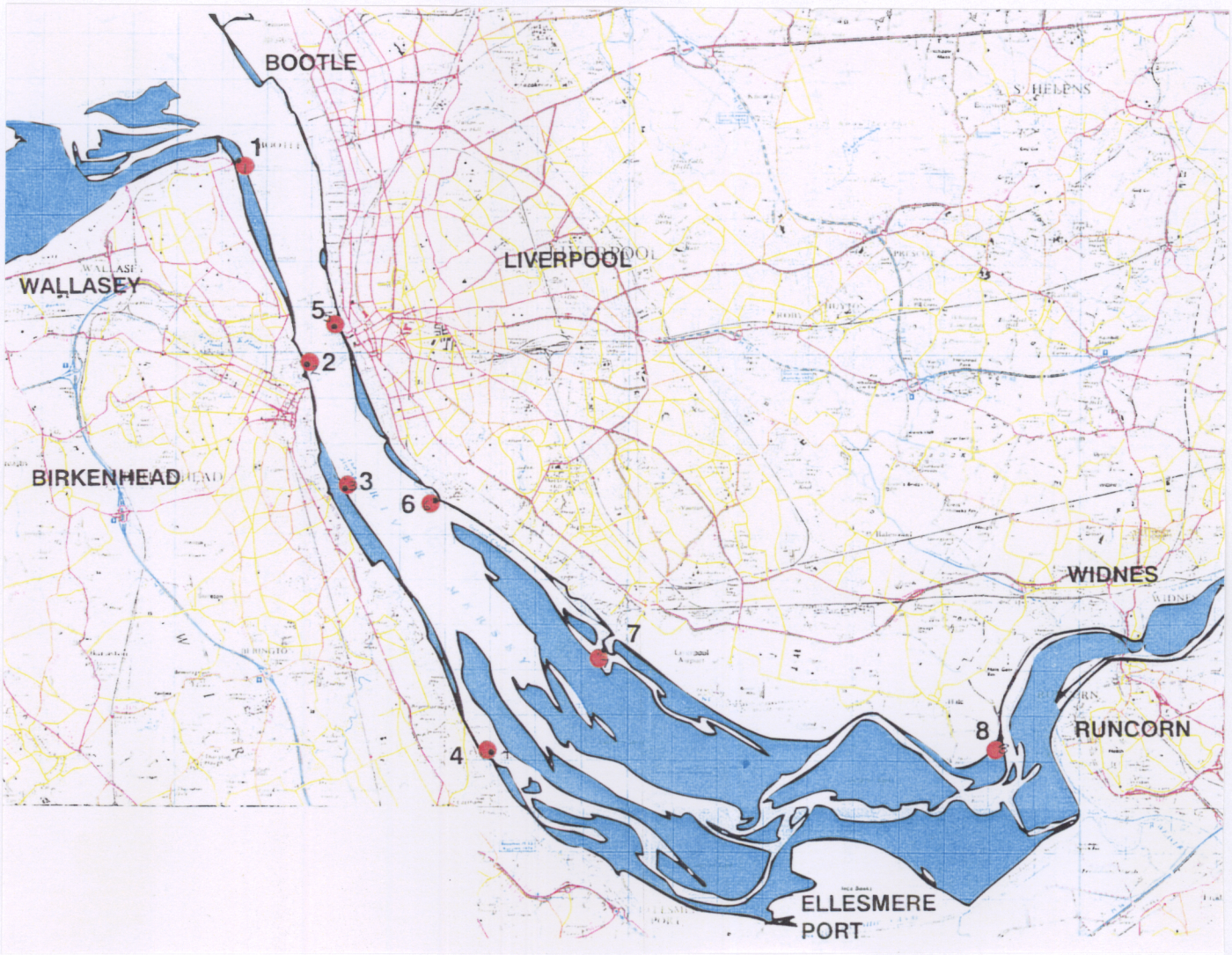

Figure 2.4 Sampling/deployment stations on the Mersey Estuary

Table 2.2 provides information on the nature of the River Aire and Willow Brook at the sampling/deployment stations in terms of slope, river width and average depth. From the data it is evident that the River Aire increases markedly in depth and width below Station 6. On the Willow Brook both width and depth increase when the three branches have joined by Stations 7 and 8.

Table 2.2 Summary information on the nature (slope, width and average depth) of the stations at the freshwater sites

Site and parameters	Station							
	1	2	3	4	5	6	7	8
River Aire								
Slope(m km^{-1})	6	0.9	ND	1.0	ND	1.0	0.6	1.1
Width (m)	6	20	ND	17	ND	15	30	30
Average depth (cm)	18.3	20	ND	43.3	ND	30	233	217
Willow Brook								
Width (m)	<1	3	0.5	2	2	5	4	8
Average depth (cm)	30	100	20	40	50	100	50	50

2.2 Study design

2.2.1 Sampling/deployment strategy

The basis on which methods were selected for use in this field evaluation has been summarised in Section 1.3 of this report and is described in greater detail in the report for Phase 2 - Stage 1 of the project (Environment Agency 1996).

The laboratory-based and *in situ* bioassays were divided into core and non-core methods (see Tables 1.4 and 1.5) depending on their developmental status (that is, those tests which were considered to require further research were classified as non-core tests) and preferences of the Environment Agency and SNIFFER. Both acute and chronic methods were included and where possible, sub-lethal measurements were included in the test design in order to improve the potential discrimination between sites of low to medium toxicity. In the freshwater 10 d *D.Magna* reproduction test and the estuarine 9 d *T.battagliai* reproduction test measurements of lethality were made after 48 h to assess the usefulness of these standardised acute procedures.

Water samples were taken for laboratory tests on a minimum of three occasions during the sampling window while sediment samples were taken on two occasions. Although this decision was based to a degree on the availability of resources it also reflected the fact that temporal variability in the sediment was likely to be greater in the water column (Rand 1996). *In situ* tests were deployed twice to coincide with sediment sampling. Reduced temporal variability was expected with these methods compared to laboratory-based methods as they

are exposed to 'in-stream' conditions for the duration of the test as opposed to those pertaining when the 'spot' water or sediment samples were taken. Table 2.3 shows the basic monitoring programme and frequency of sampling at each site.

Table 2.3 Summary of the monitoring programme

Site	No. of sampling or deployment stations	Frequency of sampling for laboratory bioassays or deployments of *in situ* bioassays			Types of bioassays used
		Water	Sediment	*In situ*	
River Aire	8	4	2	2	Core + non core
Willow Brook	8	3	2	2	Core
Tees Estuary	8	4	2	2	Core + non core
Mersey Estuary	8	3	2	2	Core

In order to assess spatial variability and to optimise statistical power of the study design, five replicate water and sediment samples were taken from each station on each sampling occasion. However, this number of station replicates would potentially be costly on a routine basis and it is possible that composite samples may adequately reflect the results of tests performed on discrete samples. Therefore, in this study, composite samples were tested in parallel with the five discrete samples (see Section 5.2).

Tables 2.4 to 2.7 show the water and sediment laboratory and *in situ* bioassays performed at each site on each occasion. These also include the number of bioassay replicates or *in situ* deployment chambers used for each sample in every test. The level of replication used was based on the guidance given in the existing protocols for the methods (see Project Record P2/i701/2).

Table 2.4 Summary of the monitoring approach used at the River Aire site

Occasion	Bioassays used	No. of stations sampled or used for deployments	No. of samples taken per station/ No. of chambers per deployment	No. of replicates per sample
1	**ECL (L-WC)**	8	5 samples (+ composite)	4
	Microtox -acute (L-WC)	8	"	4
	7 d *L. minor* growth inhibition (L-WC)	8	"	3
	10 d *D. magna* reproduction/ lethality (L-WC)	8	"	5
2	**ECL (L-WC)**	8	5 samples (+ composite)	4
	Microtox-acute (L-WC)	8	"	4
	7 d *L. minor* growth inhibition (L-WC)	8	"	3
	10 d *D. magna* reproduction/ lethality (L-WC)	8	"	5
	10 d *C. riparius* lethality (L-S)	8	"	3
	28 d *C. riparius* growth/ lethality (L-S)	8	"	3
	6 d *G. pulex* feeding rate *in situ* (IS-WC)	8	3 cages	-
	6 d *D. magna* lethality *in situ* (IS-WC)	8	4 cages	-
	10 d *C. riparius* lethality *in situ* (IS-S)	8	4 cages	-
3	**ECL (L-WC)**	8	5 samples (+ composite)	4
	Microtox-acute (L-WC)	8	"	4
	7 d *L. minor* growth inhibition (L-WC)	8	"	3
	10 d *D. magna* reproduction/ lethality (L-WC)	8	"	5
	10 d *C. riparius* lethality (L-S)	8	"	3
	28 d *C. riparius* growth/ lethality (L-S)	8	"	3
	6 d *G. pulex* feeding rate *in situ* (IS-WC)	8	3 cages	-
	6 d *D. magna* lethality *in situ* (IS-WC)	8	4 cages	-
	10 d *C. riparius* lethality *in situ* (IS-S)	8	4 cages	-
4	**ECL (L-WC)**	8	5 samples (+ composite)	4
	Microtox -acute (L-WC)	8	"	4
	7 d *L. minor* growth inhibition (L-WC)	8	"	3
	10 d *D. magna* reproduction/ lethality (L-WC)	8	"	5

Bioassays in bold are core tests, L-WC = Laboratory test - water column, L-S = Laboratory test - sediment, IS-WC = *In situ* water column test, IS-S = *In situ* sediment test

Table 2.5 Summary of the monitoring approach used for the Willow Brook site

Occasion	Bioassays used	No. of stations sampled or used for deployments	No. of samples taken per station/ No. of chambers per deployment	No. of replicates per sample
1	**ECL (L-WC)**	8	5 samples (+ composite)	4
	Microtox - acute (L-WC)	8	"	4
	7 d *L. minor* growth inhibition (L-WC)	8	"	3
	10 d *D. magna* reproduction/ lethality (L-WC)	8	"	5
2	**ECL (L-WC)**	8	5 samples (+ composite)	4
	Microtox - acute (L-WC)	8	"	4
	7 d *L. minor* growth inhibition (L-WC)	8	"	3
	10 d *D. magna* reproduction/ lethality (L-WC)	8	"	5
	10 d *C. riparius* lethality (L-S)	8	"	3
	6 d *G. pulex* feeding rate *in situ* (IS-WC)	8	3 cages	-
3	**ECL (L-WC)**	8	5 samples (+ composite)	4
	Microtox - acute (L-WC)	8	"	4
	7 d *L. minor* growth inhibition (L-WC)	8	"	3
	10 d *D. magna* reproduction/ lethality (L-WC)	8	"	5
	10 d *C. riparius* lethality (L-S)	8	"	3
	6 d *G. pulex* feeding rate *in situ* (IS-WC)	8	3 cages	-

Bioassays in bold are core tests, L-WC = Laboratory test - water column, L-S = Laboratory test - sediment, IS-WC = *In situ* water column test, IS-S = *In situ* sediment test

Table 2.6 Summary of the monitoring approach used at the Tees Estuary site

Occasion	Bioassays used	No. of stations sampled or used for deployments	No. of samples taken per station/ No. of chambers per deployment	No. of replicates per sample
1	**ECL (L-WC)**	8	5 samples (+ composite)	4
	Microtox - acute (L-WC)	8	"	4
	24 h Oyster (*Crassostrea gigas*) embryo-larval development (L-WC)	8	"	3
	9 d *T. battagliai* reproduction/ lethality (L-WC)	8	"	5
2	**ECL (L-WC)**	8	5 samples (+ composite)	4
	Microtox - acute (L-WC)	8	"	4
	24 h Oyster embryo-larval development (L-WC)	8	"	3
	9 d *T. battagliai* reproduction/ lethality (L-WC)	8	"	5
	10 d *C. volutator* lethality (L-S)	8	"	3
	28 d *C. volutator* growth/ lethality (L-S)	8	"	3
	10 d *A. marina* cast formation/ lethality (L-S)		"	2
	28 d *M. edulis* clearance rate *in situ* (IS-WC)	8	2 cages	-
	10 d *C. volutator* lethality *in situ* (IS-S)	8	4 cages	-
3	**ECL (L-WC)**	8	5 samples (+ composite)	4
	Microtox - acute (L-WC)	8	"	4
	24 h Oyster embryo-larval development (L-WC)	8	"	3
	9 d *T. battagliai* reproduction/ lethality (L-WC)	8	"	5
	10 d *C. volutator* lethality (L-S)	8	"	3
	28 d *C. volutator* growth/ lethality (L-S)	8	"	3
	10 d *A. marina* cast formation/ lethality (L-S)	8	"	2
	28 d *M. edulis* clearance rate *in situ* (IS-WC)	8	2 cages	-
	10 d *C. volutator* lethality *in situ* (IS-S)	8	4 cages	-
4	**ECL (L-WC)**	8	5 samples (+ composite)	4
	Microtox - acute (L-WC)	8	"	4
	24 h Oyster embryo-larval development (L-WC)	8	"	3
	9 d *T. battagliai* reproduction/ lethality (L-WC)	8	"	5
	10 d *C. volutator* lethality (LS)	8	4 cages	-

Bioassays in bold are core tests, L-WC = Laboratory test - water column, L-S = Laboratory test - sediment, IS-WC = *In situ* water column test, IS-S = *In situ* sediment test

Table 2.7 Summary of the monitoring approach used at the Mersey Estuary site

Occasion	Bioassays used	No. of stations sampled or used for deployments	No. of samples taken per station/No. of chambers per deployment	No. of replicates per sample
1	**ECL (L-WC)**	8	5 samples (+ composite)	4
	Microtox - acute (L-WC)	8	"	4
	24 h Oyster embryo-larval development (L-WC)	8	"	3
	9 d *T. battagliai* reproduction/ lethality (L-WC)	8	"	5
2	**ECL (L-WC)**	8	5 samples (+ composite)	4
	Microtox - acute (L-WC)	8	"	4
	24 h Oyster embryo-larval development (L-WC)	8	"	3
	9 d *T. battagliai* reproduction/ lethality (L-WC)	8	"	5
	10 d *C. volutator* lethality (L-S)	8	"	3
	28 d *M. edulis* clearance rate *in situ* (IS-WC)	8	2 cages	-
3	**ECL (L-WC)**	8	5 samples (+ composite)	4
	Microtox - acute (L-WC)	8	"	4
	24 h Oyster embryo-larval development (L-WC)	8	"	3
	9 d *T. battagliai* reproduction/ lethality (L-WC)	8	"	5
	10 d *C .volutator* lethality (L-S)	8	"	3
	28 d *M. edulis* clearance rate *in situ* (IS-WC)	8	2 cages	-

Bioassays in bold are core tests, L-WC = Laboratory test - water column, L-S = Laboratory test - sediment, IS-WC = *In situ* water column test, IS-S = *In situ* sediment test

2.2.2 Sampling/deployment procedures

General

All water and sediment samples were collected and transported to the laboratory within 12 hours. At each of the estuarine sites sampling was performed over a single low tide where possible. Laboratory water tests were started within 24 hours and sediment tests were performed within 48 hours. Freshwater samples were stored in five litre polythene bottles and estuarine samples in one litre polythene bottles.

Water samples

At all sites five discrete water samples were collected from bank-side or fixed structures at a sampling station using a plastic bucket (rinsed with site water prior to every sample collection) with the time delay between each sample taken being minimised. Bottles were filled to the top to exclude air. After storage overnight (in the dark at ambient), water was decanted from every sample and any solid material discarded. Water quality measurements (temperature, pH, dissolved oxygen and hardness) were taken on the bulk sample before testing commenced. The salinity of estuarine water samples was measured and salinity adjusted to 34 ppt using Aristar sea salt where necessary. An additional composite sample was prepared from the five spot samples after any manipulations, by mixing aliquots of each sample. The physico-chemical measurements are reported in detail in the accompanying Project Record P2/i701/2. Samples with low dissolved oxygen levels were aerated to >60% of the air saturated value (using an oil free compressed air supply) before being used in water column bioassays.

Sediment samples

Five discrete sediment samples were collected from depositional areas of rivers and inter-tidal estuarine mudflats. Two kilograms of freshwater sediment was collected using an Ekman grab sampler. Visually oxic surface sediment (6.5 kg) was collected from random locations on estuarine mudflats using a spade. Samples were transported in 10-litre sealed polythene buckets and stored at room temperature overnight. All the sediments were sieved through a coarse 2-mm metal sieve with site water to remove debris. Any indigenous animals observed were removed using tweezers. An additional composite sample was constructed from the five spot samples after sieving by mixing aliquots of each sample. Sub samples for particle size distribution, total organic carbon (TOC) analysis and water content determinations were taken from all discrete and composite samples after sieving.

2.2.3 Laboratory toxicity bioassays

Laboratory bioassays were performed according to existing protocols (see accompanying Project Record P2/i701/2). Table 2.8 provides summary information on the type of bioassays used, the bioassay responses and specific endpoint(s) of each and the number of organisms exposed in each test replicate. This table also indicates those methods for which a laboratory reference control was used, that is a standard dilution water as opposed to a reference site water. No reference control was used for the sediment tests. Tables 2.9 and 2.10 provide a more detailed summary of the methods including bioassay organisms and the systems and equipment used.

Table 2.8 Summary information on laboratory test methods (For further details see Project Record P2/i701/2)

Bioassay (duration)	Type of bioassay	Bioassay response	Bioassay endpoint	Reference control included	Number of organisms used per bioassay replicate
ECL (4 min)	Laboratory, water column, freshwater and estuarine	Light production in chemiluminescent method	ECL Units	Yes	NA
Microtox-acute (5-30 min)	"	Light production in bioluminescent method	% response	"	NA
L. minor growth (7 d)	"	Growth inhibition	Growth rate	"	20
D. magna reproduction (10 d)	Laboratory, water column, freshwater	Juvenile production, lethality after 2 d and 10 d	Numbers of juveniles, % lethality	"	5
Oyster embryo-larval development (24 h)	"	Embryo development	Percent Net Response (PNR)	"	1500
T. battagliai reproduction (9 d)	Laboratory, water column, estuarine	Juvenile production, lethality after 2d and 9 d	Numbers of juveniles, % lethality	"	5
C. riparius lethality (10 d)	Laboratory, sediment, freshwater	Lethality	% lethality	No	20
C. riparius growth (28d)	Laboratory, sediment, freshwater	Growth, lethality	Dry weight, % lethality	No	20
C. volutator lethality (10 d)	Laboratory, sediment, estuarine	Lethality	% lethality	No	20
C. volutator growth (28d)	Laboratory, sediment, estuarine	Growth, lethality	Dry weight, % lethality	No	20
A. marina lethality (10 d)	Laboratory, sediment, estuarine	Cast formation, lethality	Number of casts, % lethality	No	2

Table 2.9 Summary of organisms and equipment needed for the laboratory-based bioassays on freshwater water column and sediment samples (for further details see Project Record P2/i701/2)

Parameter	Freshwater laboratory-based bioassays					
	ECL	Microtox -acute	*Lemna* growth inhibition	*Daphnia* reproduction	*Chironomus* lethality	*Chironomus* growth/lethality
Age of organisms	Not applicable	Not applicable	Not appropriate	<24 hr old neonates	2nd instars (10-12 days old)	Egg packets
Test temperature	20 °C	15 °C	20 °C	20 °C	20 °C	20 °C
No. of samples	48	48	48	48	48	48
No. of replicates per sample	4	4	3	5	3	2 (elutriate) 4 (sediment)
Total no. of vessels	192	196	144	240	144	96 (elutriate) 192 (sediment)
Type of vessels	Cuvettes	Cuvettes	Crystallizing dishes	125 ml beakers	250 ml beakers	Crystallising dishes (elutriate), 100 ml beakers (sediment)
No. of control vessels	1	5	10	10	3	2 elutriate/4 sediment
Reference toxicant	Phenol	Zinc (as Zn SO_4)	Zinc (as Zn SO_4)	Zinc (as Zn SO_4)	-	-
No. of concentrations	5	5	5	5	-	-
No. of replicates	1	1	3	5	-	-
Total no. of vessels	6	5	15	25	-	-
No. of organisms per vessel	Not applicable	Not applicable	20 fronds	1	20	1 egg packet (~200 organisms)
Total no. of organisms required	Not applicable	Not applicable	3380	275	2940	98 egg packets
Materials required	Reagents 1,2 and 3, HPLC water	Reagent (3 vials), Media (2 vials)	169x 110 ml crystallising dishes	550x125 ml (tall form) beakers	147x250 ml beakers, 16 plenum chambers	98 crystallising dishes, 147x100 ml beakers, 16 plenum chambers

Table 2.10 Summary of organisms and equipment needed for the laboratory-based bioassays on estuarine water column and sediment samples (for further details see Project Record P2/i701/2)

Parameter	Estuarine laboratory-based bioassays						
	ECL	Microtox - acute	Oyster embryo-larval development	*Tisbe* reproduction	*Corophium* lethality	*Corophium* growth/lethality	*Arenicola* cast formation
Age of organisms	Not applicable	Not applicable	Oyster embryos	11 day old adults	Adults	Juveniles (600-700 µm sieve)	Adults (1-2 g)
Test temperature	20 °C	15 °C	25 °C	20 °C	15 °C	15 °C	15 °C
No. of samples	48	48	48	48	48	48	48
No. of replicates per sample	4	4	4	10	3	3	3
Total no. of vessels	192	196	192	480	144	144	144
Type of vessels	Cuvettes	Cuvettes	35 ml glass pots	12 well culture plates	1000 ml beakers	1000 ml beakers	2000 ml beakers
No. of control vessels	1	5	6	10	3	3	3
Reference toxicant	Phenol	Zinc (as Zn SO_4)	Zinc (as Zn SO_4)	Zinc (as Zn SO_4)	-	-	-
No. of concentrations	5	5	5	5	-	-	-
No. of replicates	1	1	3	10	-	-	-
Total no. of vessels	6	5	15	50	-	-	-
No. of organisms per vessel	Not applicable	Not applicable	1500	1	20	20	5
Total no. of organisms required	Not applicable	Not applicable	319500	540	2940	2940	735
Materials required	Reagents 1,2 and 3, HPLC water	Reagent (3 vials), Media (2 vials)	5 male + 5 female oysters, 213 glass pots	54x12 well culture plates	147x1000 ml beakers, 16 plenum chambers	147x1000 ml beakers, 16 plenum chambers	147x2000 ml beakers, 16 plenum chambers

2.2.4 *In situ* toxicity bioassays

Table 2.11 gives summary information of the *in situ* methods used and Table 2.12 gives additional experimental detail of the organisms and equipment used. The animals were transported to the study site on the day of deployment and acclimated to field temperatures in transit. The temperature of site water was measured prior to deployment.

Daphnia, Gammarus and *Chironomus* cages were staked to the bottom of river beds in close proximity to the sediments but in positions least likely to be affected by vandalism. Mussels (*Mytilus edulis)* were deployed from fixed structures or buoys as necessary. *Corophium* deployments were situated on mud flats at low tide, but in positions where they could be collected at low tide 10 days later. The *Daphnia magna* juveniles (<24 h old) and the 10 d old *Chironomus riparius* larvae were obtained from cultures at WRc. The *Gammarus pulex* and *Corophium volutator* were obtained from clean locations on the River Darenth (Kent) and in the Netherlands. The mussels were obtained from Guernsey Sea Farms.

Table 2.11 Summary information on *in situ* bioassays

Bioassays (duration)	Bioassay response	Bioassay endpoint	Number of organisms used per *in situ* cage
D. magna lethality (6 d)	Lethality	% lethality	10
G. pulex feeding rate (6 d)	Consumption of leaf material, lethality	Feeding rate, % lethality	16
M. edulis clearance rate (28 d)	Algal clearance	Clearance rate, % lethality	30
C. riparius lethality (10 d)	Lethality	% lethality	10
C. volutator lethality (10 d)	Lethality	% lethality	10

Table 2.12 Summary of organisms and equipment needed for *in situ* deployments

Parameter	*In situ* deployments				
	Gammarus pulex	*Daphnia magna*	*Chironomus riparius*	*Mytilus edulis*	*Corophium volutator*
Age of organisms	Adults	<24 h old neonates	2nd instars (10-12 day old)	Adults (4.5-5 cm)	Adults <8 mm
Deployment duration	6 days	6 days	10 days	28 days	10 days
No. of deployment stations per site	8	8	8	8	8
Type of holding vessel	Cages (Holding baskets)	Cages	Cages	Cages	Cages
No. of replicate vessels per station	50 (in 3 holding baskets)	4	4	2	4
No. of control vessels station	16 leaf controls (in 3 holding baskets)	-	-	-	-
No of organisms per holding vessel	1	10	10	30	10
Total no. of organisms required per station	50	50	50	60	50
Total no. of organisms required per site	400	320	320	480	320
Materials required	528 cages, 24 holding baskets, 60x Sterilin trays	32 cages, 8 stakes	32 cages, 16 stakes	16 cages	32 cages, 32 stakes

2.2.5 Biomarker studies

As part of the overall package of ecotoxicological methods used, the University of Plymouth carried out a series of biomarker studies on the *Mytilus edulis* deployed at the Tees and Mersey Estuary sites.

Test organisms

Metallothionein: On the Tees Estuary 64 mussels (*Mytilus edulis*) which had been deployed at eight separate stations (eight mussels per station; four stations on either side of the estuary) between 30 August - 27 September 1996 (mussels from Station 5 were not collected because of tidal conditions) as spares for clearance rate measurements were collected. Similarly, 64 mussels deployed along the Mersey Estuary between 25 September - 23 October 1996 (mussels from Station 8 were not collected because of tidal conditions) were collected. Following collection, the mussels were dissected (on the same date as collection) and the whole tissues immediately frozen in liquid nitrogen. These samples were stored at -80 °C until metallothionein (MT) analysis.

Lysosomal integrity (neutral red) and Cardiac Activity: One hundred and thirty six mussels (shell length 4.15-4.87 cm) which had been deployed along the Mersey estuary (same stations as MT) between 28 August - 25 September 1996 were used. Similarly, 136 mussels (shell length 4.11-5.19 cm) which had been deployed along the Tees Estuary (same stations as MT) between 27 September - 25 October 1996 were used. Nine animals were used per station for the neutral red assay and eight animals per station for heart-rate measurements. Following deployment, the mussels were transferred to the marine aquarium at the University of Plymouth (on the same day as collection under cool conditions) and maintained in aerated seawater from their respective stations at 15 °C in 2 litre tanks (mussels from Station 3 on the Tees were not collected because of tidal conditions). The animals were not fed during a one day holding period or during subsequent experiments.

Metallothionein analysis

Quantification of MT was determined by partial purification as described by Pederson *et al* (1994) with some modifications. Frozen (liquid nitrogen) whole tissue was placed in a mortar, filled with liquid nitrogen and ground to a fine consistency. Approximately 1 g of sample was transferred to a 10 ml beaker, cold distilled water containing 1 mM DTT (dithiothreitol) and 0.1 mM PMSF (phenylmethylsulfonylfluoride) was added. The suspension was sonicated and centrifuged for 70 min at 55 000 rpm at 4 °C.

The supernatant was prepared for a direct method of MT quantification (using spectrophotometry) as described by Dieter *et al* (1987) and Viarengo *et al* (1995). A two-step ethanol/chloroform precipitation of the supernatant (with centrifugation) was performed and the resulting pellet was resuspended in 5 mM Tris-HCl, 1 mM EDTA buffer, pH 7, followed by reaction with DTNB (5,5 dithiobis 2 nitrobenzoic acid) as described by Viarengo *et al.* (1995). Sample absorbance was read at 412 nm in a UNICAM UV/VIS spectrophotometer, and the MT concentration was quantified using glutathione (GSH) as a reference standard (Viarengo *et al* 1995). The Statgraphics 5.0 software package was used to examine differences between stations using one-way ANOVA and multiple range tests of least significant difference (LSD). Bartletts test was applied to check the homogeneity of variances.

Neutral red assay

Haemocytes were withdrawn from the posterior adductor muscle of the mussel, using a hypodermic syringe, and placed into physiological saline solutions (HEPES (N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulphonic acid]) 4.77 g, NaCl 25.48 g, $MgSO_4$

13.06 g, KCl 0.75 g and $CaCl_2$ 1.47 g l^{-1}; pH 7.4). The haemocytes were then allowed to adhere to poly-L-lysine treated slides for 30 minutes. Excess solution was then removed from the slides and the remaining monolayer treated with neutral red dye (neutral red stock solution: 20 mg ml^{-1} DMSO (Dimethyl sulphoxide), working solution 5 $\mu g\, ml^{-1}$ in physiological saline). The cells were incubated for a further 15 minutes. After the incubation with neutral red, the slides were examined (microscopically) immediately, 15 minutes later and then at 30 minute intervals thereafter. Cells were assessed for leakage of the dye from the lysosomes into the cytosol. When 50% of the cells on a slide had reached this stage the endpoint was considered to have been attained and the maximum retention time (in mins) was recorded. Stratigraphics 5.0 was used to examine differences between sites using one-way ANOVA and multiple range tests (LSD). Bartletts test was applied to check the homogeneity of variances.

Cardiac activity

The mean resting heart rate of each individual was measured over seven hours using the technique described in detail by Depledge and Andersen (1990). This recording system permitted 16 animals to be monitored simultaneously. Infra-red transducers were glued (using Loctite 314 adhesive) directly to the shell, on the mid dorsal line just behind the posterior termination of the hinge. The infra-red transducers were used to monitor the heart's cycle of action and heart rate was determined every minute and stored on computer floppy disk for later analysis. Care was taken to ensure that the mussels were not disturbed during the recording periods.

Bartlett's variance check indicated that mean heart rates of individuals were not normally distributed (because the heart stops beating periodically). Therefore, distribution free statistics were used (Krustall-Wallis test). Data are shown as Box and Whisker plots. The box was divided at the median. Fifty percent of the data points are encompassed by the length of the box. The vertical bars drawn from the top and bottom of the box include 95% of the data points.

2.2.6 Physico-chemical parameters

A series of physico-chemical determinands (particle size, organic carbon, water content) were measured in each sediment sample to aid in the interpretation of the resulting toxicity data. These parameters are known to have the potential to influence the results from sediment toxicity tests (SETAC 1993) and the data may need to be normalised on the basis of differences in the physico-chemical parameters between locations.

Particle size analysis

A representative sub-sample was wet sieved through a series of sieves with mesh diameters of 2 mm, 1 mm and 0.5 mm respectively. The retained fraction on each screen and the proportion passing through the 0.5 mm sieve was dried to constant weight at 60 °C. The dry weight of each fraction was recorded. A representative sub-sample of the <0.5 mm fraction was resuspended in water and analysed using laser defraction granulometry. The distributions of the particle size in the >0.5 mm fraction, determined by wet sieving, and the <0.5 mm

fraction, determined by laser defraction granulometry, were combined into a single distribution. The results were expressed as percentage composition of each sample in grades of particle size based on the Wentworth scale.

Organic carbon

Analyses for these determinands were undertaken on the <63 μm fraction of the sediment. Each sample was freeze dried and sieved through a nest of stainless steel sieves culminating in a sieve with a mesh diameter of 63 μm. The retained fraction remaining on the sieves was discarded and the fraction passing through the 63 μm sieve was weighed, recorded and retained for the analysis of total organic carbon (TOC) and total nitrogen (TN).

Determination of TOC required compensation for the presence of carbonates. The method used required the prepared sample to be sub-sampled. One portion was combusted in a muffle furnace at 475 °C for 4 hours to remove organic carbon but not total inorganic carbon (TIC) (that is carbonates). The other portion was not combusted. The combusted sub-sample was weighed before and after combustion and the weight loss recorded as percentage loss-on-ignition (LOI). Both combusted and uncombusted samples were analysed for elemental carbon using a CHN elemental analyser. TOC was calculated by subtracting total carbon (TC) in the combusted sample, where TC = TIN, from TC in the uncombusted sample where TC = TIN + TOC. TOC was expressed as a percentage of the weight of the original sample.

TN was determined on the uncombusted sample using the CHN elemental analyser and expressed as a percentage of the original sample weight.

Water content

A sub-sample of each sediment was weighed and the then dried at 105 °C to constant weight. The water content was then calculated.

2.2.7 Macroinvertebrate surveys

Field methods

River Aire

Samples on the River Aire were taken on 8 September 1996 by the Environment Agency North East Region at six of the eight stations used for the ecotoxicological techniques (Table 2.1), the exceptions being Stations 3 (Downstream of Marley STW) where there were no riffles and Station 7 (Castleford) where the water depth precluded sampling (see Table 2.2). A three minute kick sample (Stations 1,2,4,5,6) or airlift sample (Station 8) was taken at each station. Observations on site were also made of habitat characteristics such as the substrate type (boulders, cobbles, pebbles, gravel, sand and silt) and cover (algae, moss, macrophytes), river depth, flow and width and conductivity.

Willow Brook

Samples on the Willow Brook were taken on 20 September 1996 by the Environment Agency Anglian Region at all eight stations used for the ecotoxicological techniques (Table 2.1). A three-minute kick sample was taken at each station.

Tees Estuary

Samples on the Tees Estuary were taken between 28-30 October 1996 by the Environment Agency North East Region at six of the eight stations used for the ecotoxicological techniques (see Table 2.1), the exceptions being Stations 3 (Upstream of Dabholm Gut) and 6 (Downstream of Dabholm Gut) where there was no soft sediment at the time of sampling. Five replicate core samples (area 0.1 m^2) were taken at each of the six sites. At each site one sediment sample was taken for particle size analysis (PSA) and organic carbon measurements.

Mersey Estuary

Samples on the Mersey Estuary were initially planned to be taken by WRc at all of the eight stations used for the toxicological techniques (see Table 2.1). On 1-2 October 1996 four replicate samples were taken at each station using a corer (internal diameter 10 cm). However, limited numbers of organisms were found in most samples (see Project Record P2/i701/2 for further details).

Consequently it was decided to repeat the survey and to sample sediment that appeared to support a fauna based on field observations. There were insufficient resources to repeat sampling at the original eight stations but four of the stations offered the opportunity for sampling a standard sediment type and these represented a gradient of biological quality.

On 22-23 October 1996 sampling was repeated at Stations 3 (Tranmere) and 4 (Eastham Locks) on the Wirral bank and Stations 6 (Dingle) and 8 (Hale) on the Liverpool bank. Four replicate samples were taken using a larger corer (internal diameter 14 cm) in the same way as in the previous survey. Samples for PSA were taken at each site in the same way as in the previous survey but interstitial salinity was not measured on this occasion.

Laboratory methods

Freshwater samples

The benthic macro-invertebrates present were identified to family level by Environment Agency biologists and the numbers present recorded. The data were used to determine for each station the number of taxa present, the Biological Monitoring Working Party (BMWP) score and the Average Score Per Taxa (ASPT). The number of organisms in the different taxa were recorded for samples from the River Aire whereas for the Willow Brook estimates were made of the numbers according to a series of bands (present = 1 specimen, scarce = 2-10, common = 11-100, abundant = 101-1000 and very abundant = 1000+).

Estuarine samples

Each sample was rinsed thoroughly in freshwater to remove the formalin fixative. The samples were examined using a low power stereo microscope and all animals were identified to species level, where possible, and preserved in 70% industrial methylated spirit (IMS). The taxonomy was standardised using the Species Directory to British Marine Fauna and Flora (Howson 1987). A reference collection has been retained for future use.

The number of animals in each taxa was enumerated using the convention that the presence of an individual was recorded positively if the whole animal or the anterior portion was present.

Biomass determinations were made on a blotted wet weight basis. After the three month stabilisation period in fixative the animals were blotted dry to remove excess preservative. Bivalve molluscs (for example cockles) and echinoderms (for example urchins) were opened or punctured sufficiently to release preservative in the body cavities. The minimum weight measured was standardised to 0.1 mg.

2.3 Data analysis

2.3.1 Introduction

The data analysis has been carried out in the following way:

1. Analysis of bioassay data including:

    - three and two way analysis of variance (ANOVA) to determine the pattern of response (that is the variability due to large scale station and small scale sample/cage spatial effects and also temporal effects) for each bioassay at the different sites (see Section 3.1);

    - an assessment of the correlations between the responses of bioassays used at different sites to investigate redundancy of information, that is whether different bioassays are providing the same information (see Section 3.2);

    - an initial assessment of the responses of each method at the reference location and positive control stations on the River Aire and Tees Estuary sites using non-parametric techniques (such as Kruskal-Wallis and Mann-Whitney tests). These stations represent the extremes of the chemical contamination gradient and for tests to be useful they need to be able to discriminate between these extremes (see Section 3.3);

    - confirmatory analysis of the discrimination of each method using all the available data from the study sites. The analysis involved comparing responses at each station with those at the relevant reference location using the data from the three and two-way ANOVA and Dunnetts tests (see Section 3.3);

- an assessment of the power of the bioassays based on the test design used in the study to assess the level of difference in response that a bioassay can detect (see Section 3.4).

2. Comparison of the ecotoxicological data with analytical chemistry and macro-invertebrate data including:

   - analysis of biological data using univariate and multivariate measures of community composition to determine the pattern of response between stations at the different sites (see Section 3.5);

   - an assessment of the extent to which bioassay responses for bioassays used at stations on the River Aire and Tees Estuary are correlated with the analytical chemistry on a suite of known pollutants and macroinvertebrate survey data (see Section 3.5);

   - an assessment of the extent to which the relationships between bioassay data macroinvertebrate data identified for the River Aire and Tees Estuary sites are applicable to the Willow Brook and Mersey Estuary sites (see Section 3.5).

2.3.2 Analysis of bioassay data

Assessment of the patterns of response of bioassay data

At each site the data for each bioassay were initially assessed using three factor analysis of variance with sample, station and time as the factors. For each bioassay the responses were transformed, where necessary, to obtain a normally distributed dataset. Table 2.13 shows the transformations used for each method. Square root transformations were used where group variances were proportional to the means (the ECL bioassay), while arcsine transformations were used for percentage lethality data (the *D. magna* and *T. battagliai* reproduction bioassays, the 10 and 28 day *C. riparius*, *C. volutator* and *A. marina* sediment bioassays and the *D. magna*, *G. pulex*, *C. riparius* and *C. volutator in situ* bioassays).

If the 3-way ANOVA showed there was temporal variability for a bioassay then 2-way ANOVA was carried out on the responses on each sampling/deployment occasion. Analysis was carried out using Genstat 5, version 3.1 software (Rothampstead Experimental Station). For the *G. pulex* feeding rate bioassay ANOVA was carried out on the leaf correction factors and *G. pulex* animal weight to determine whether these factors influenced feeding rate responses at stations

Initial assessment of the differences in bioassay responses

The comparison of the data obtained for each bioassay at the reference location and positive control station at the River Aire and Tees Estuary sites was carried out using non-parametric techniques such as Kruskal-Wallis or Mann-Whitney tests (Zar 1984). Non-parametric techniques were used because the responses from the reference control and positive control stations were generally bimodally, rather than normally, distributed representing the two extremes (no or low effect and complete effect) of the response range.

Table 2.13 Transformations used in the analysis of variance of bioassay data

Bioassay	Transformation used
ECL	Square root transformation
Microtox - acute	No transformation
L. minor growth inhibition	"
D. magna reproduction - juvenile production	"
D. magna reproduction - lethality	Arcsine transformation
G. pulex in situ - feeding rate	No transformation
G. pulex in situ - lethality	Arcsine transformation
C. riparius lethality	"
D. magna in situ - lethality	"
C. riparius in situ - lethality	"
Oyster embryo-larval development	Arcsine transformation
T. battagliai reproduction - juvenile production	No transformation
T. battagliai reproduction - lethality	Arcsine transformation
M. edulis in situ - clearance rate	No transformation
C. volutator lethality	Arcsine transformation
A. marina cast formation	"
A. marina lethality	"
C. volutator in situ - lethality	"

Confirmatory analysis of the discrimination of each bioassay

In the confirmatory analysis of the data from each bioassay the data from the three and two ANOVA were used along with Dunnett's tests to identify for each bioassay which stations at a site showed statistically significant differences from responses at the relevant reference location.

Assessment of the power of the bioassays

In the assessment of the power of each bioassay the minimum detectable difference (MDD) which each bioassay was capable of detecting at each site was calculated using the following formula (Zar 1984):

$$\text{MDD } \delta = \sqrt{(2ks^2\phi^2)/n'}$$

where:

k = number of stations

s^2 = within cells mean square (MS) from the relevant 2-way ANOVA

ϕ^2 = power of the test obtained from curves in Zar (1984) using $\alpha = 0.05$, $\beta = 0.1$, v_1 = factor degrees of freedom from the ANOVA, v_2 = within cells (error) degrees of freedom from the ANOVA. The α and β values used represent typical levels used to assess the power of bioassays (Zar 1984) as described in Section 3.4.

n' = number of values

2.3.3 Comparison of the ecotoxicological data with analytical chemistry and macroinvertebrate survey data

Analysis of biological survey data

The status of macrobenthic communities can be described in a number of ways; primary community attributes, derived measures involving a combination of primary attributes, summary measures describing diversity or pollution status, and finally multivariate representations of community composition. In freshwaters, the commonly used variables are the number of identified taxa (at the family level), and two biotic indices: the Biological Monitoring Working Party (BMWP) score and the Average Score Per Taxon (ASPT) score. Multivariate representations of community composition are not widely used but offer a further expression of community structure. In estuarine and marine waters, examples from all categories of measures are commonly used.

Primary and derived variables

The primary community attributes are: the total number of identified and quantified taxa (T), the total number of individuals, or abundance (A) and the total biomass (B). From these primary variables the derived variables, the ratios A:T (the average number of individuals per taxon - a measure of dominance), and B:A (the average weight of the animals in a sample) were calculated.

Diversity indices

Measures of diversity attempt to integrate two aspects of the community structure, namely the number of taxa (species richness) and the distribution of individuals among those taxa (evenness or equitability). A wide range of indices commonly used to describe diversity including Shannon-Wiener (H') and Pielou's index of evenness (J') and Simpson's diversity index (D), usually expressed as 1/D, have been developed and these were calculated for the data in this study.

Multivariate techniques

Data arising from studies of ecological communities have some unique characteristics: there are many variables (species), there is a predominance of zero values and the distribution of the counts or weights is usually highly right skewed. These characteristics preclude the use of almost all conventional statistical techniques on the raw data. Only when summary statistics or indices are calculated can these techniques be employed. Multivariate statistical techniques can deal with large numbers of variables and the problems of many zero values and skewed distributions can be overcome by the choice of the appropriate technique and transformation of the data.

Multivariate techniques can be broadly classified into ordination and classification methods. Ordination methods are used most commonly in the analysis of macrobenthic community data and of the many techniques available the two most commonly used for this purpose are: Detrended Correspondence Analysis (DECORANA) and non-metric multidimensional scaling

(MDS) based on Bray-Curtis similarity. The non-parametric nature of the latter techniques offers certain advantages over the former and has been used in the analyses of these data.

Comparison of ecotoxicological data, analytical chemistry and macroinvertebrate survey data

The approach taken in the comparative analysis of the ecotoxicological and the biological survey data from the River Aire and Tees Estuary sites has been to determine the pattern of response in the biological data and then, by various methods, to ascertain the degree to which the ecotoxicological data reflects the pattern of response in the biological data. By using the response in the biological data as the point of reference, there is an implicit assumption that this response is the "definitive" response and that the performance of the ecotoxicological data can be measured with reference to this. In reality, there will be scenarios where the responses of the biological and ecotoxicological data might be expected to give similar responses, for example in very contaminated conditions (Birge *et al* 1989, Eagleson *et al* 1990, Dickson *et al* 1992) such as exist at the River Aire and the Tees Estuary. There will also be scenarios when the response of the biological and ecotoxicological data might be expected to be different, for example in conditions of moderate to low contamination (Dickson *et al* 1992). In these situations, the extra sensitivity of the ecotoxicological response might be expected to give a signal but the effects might be such that they have not become apparent at the community level and, therefore, a signal in the biological data might not be present. Across the spectrum of contamination, the biological and ecotoxicological responses could provide complementary information which together provide a better picture of environmental quality (see Section 1.2).

A Bray Curtis similarity matrix was constructed from the Aire and Tees macroinvertebrate abundance data after 4^{th} root transformations to reduce the influence on the analyses of the more abundant taxa. These data were then plotted as hierarchical dendrograms, using group average linkage.

Water chemistry data were transformed [log (x+1)] and related to the macroinvertebrate data using weighted Spearman's rank correlation coefficients (Clarke and Ainsworth 1993) implemented in the PRIMER software programme (Clarke and Warwick 1994). Chemical data from the Aire comprised pH, BOD5, hardness, and concentrations of chloride, solids. dissolved oxygen, unionised ammonia, copper and zinc. Physico-chemical data from the Tees comprised percentage silt in the sediment, and concentrations of cadmium, chromium, copper, lead, mercury, nickel, zinc. lindane, dieldrin, hexachlorobenzene and polychlorinated biphenyls in the sediment. This information was not collected as part of the study but was the information available from local regulators as part of their routine monitoring during the period of the study.

Bioassay data from both sample collection/deployment dates in the Aire and the Tees were averaged, transformed [square root (x+1)], and then related to the macroinvertebrate data using weighted Spearman's rank correlation coefficients. The same procedure was used to relate water chemistry data to bioassay results.

3. OVERALL ASSESSMENT OF THE BIOASSAYS AND BIOMARKERS

This section provides an overall assessment of the bioassay and biomarker methods considering the sensitivity, discrimination, power, cost and ecological relevance (that is relationship with biological survey data) of the methods. Raw data for each bioassay is given in Project Record P2/i701/2).

3.1 Summary of the results for each method

In the initial consideration of each method, the data has been represented graphically and the responses at each station have been ranked on each occasion from least to most impacted. Three way analysis of variance (ANOVA) has been carried out on each dataset to assess the extent of spatial (between samples and stations) and temporal (between sampling occasions) variability. If this analysis showed there was temporal variability two-way ANOVA (station and sample effects) was carried out using the data for each sampling/deployment occasion.

Table 3.1 summarises the data for physico-chemical parameters (dissolved oxygen, pH and hardness or salinity) measured in water samples taken at each site. It would appear that the physico-chemical conditions at the freshwater sites at the time of sampling would not have been sufficient in themselves to result in any effects in the bioassays. For the estuarine water samples the physico-chemical parameters indicated that salinity correction and aeration of samples from certain stations was required prior to testing.

Table 3.1 Summary of the physico-chemical data for water samples taken at the different stations at each site during the study period

Site and parameters	Range of values at different stations							
	1	2	3	4	5	6	7	8
River Aire								
DO (% ASV)	94-96	35-100	60-100	67-100	67-89	67-91	68-90	55-66
pH	8.0-8.7	7.6-8.2	7.5-7.8	7.6-77	7.3-7.6	7.4-7.6	7.3-7.5	7.2-7.5
Hardness	168-	172-	-	146-	153-165	-	174-	206-
(mg $CaC0_3$/l)	187	186		159			227	228
Willow Brook								
DO (% ASV)	>60	>60	>60	>60	>60	>60	>60	>60
pH	8.4-8.9	7.4-7.5	7.5-7.7	7.6-7.7	7.6-7.7	7.8-8.1	8.4-8.8	7.8-7.9
Tees Estuary								
DO (% ASV)	<60	<60	<60	<60	<60	<60	>60	>60
pH	7.0-7.7	7.2-7.6	7.2-8.5	7.3-7.4	7.8-8.5	7.4-7.9	7.1-7.7	7.3-7.9
Salinity (‰)	25-35	22-35	16-35	23-34	12-31	22-31	26-35	32-35
Mersey Estuary								
DO (% ASV)	>60	>60	>60	>60	>60	>60	>60	<60
pH	7.8-7.9	7.2-7.5	7.1-7.5	7.2-7.5	7.1-7.5	7.1-7.5	7.0-7.4	7.0-7.3
Salinity (‰)	32-34	25-28	26-30	24-26	21-31	18-27	18-25	0-11

3.1.1 Freshwater toxicity bioassays - water column

Enhanced chemiluminescence (ECL) bioassay

Figures 3.1 and 3.2 show the results of ECL bioassays performed on four occasions for the River Aire and three occasions for the Willow Brook. Water samples were collected at all eight stations at each site on each occasion.

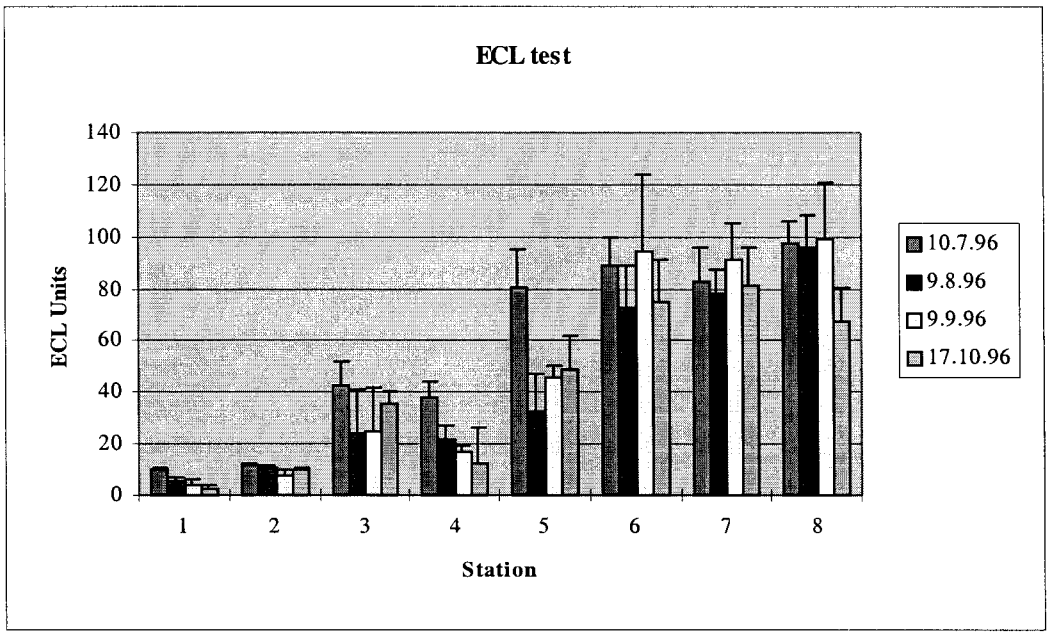

Figure 3.1 Mean ECL responses on four sampling occasions for the River Aire (error bars represent one standard deviation of station samples from the mean)

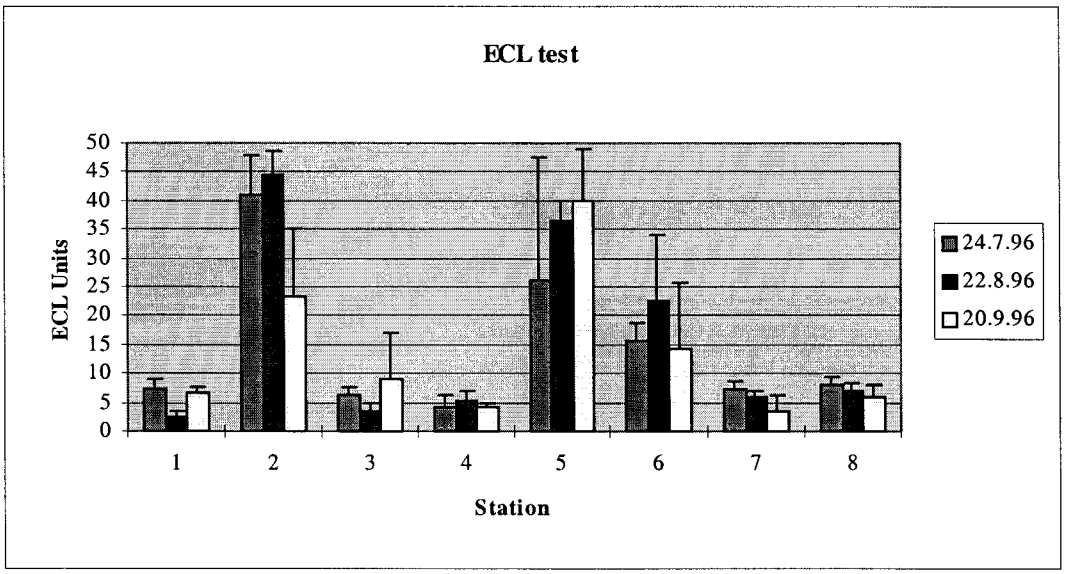

Figure 3.2 Mean ECL responses three sampling occasions for the Willow Brook (error bars represent one standard deviation of station samples from the mean)

Table 3.2 shows the results of the three and two way ANOVA and the ranking of responses between stations. ANOVA demonstrated that there were statistically significant differences in ECL response between stations at both the River Aire and the Willow Brook. This was the case when the sampling occasions were considered together and separately (see Table 3.2).

The 3-way ANOVA showed the differences in mean ECL responses with sampling date to be statistically significant for both rivers (River Aire: variance ratio = 23.9, $p<0.001$, df=3, Willow Brook: variance ratio = 3.46, $p<0.05$, df=2) respectively. This may have been due to flow conditions up to and including the sampling time and patterns in the discharge of contaminants. In addition, an element of spatial variability may be included in the temporal variability as exactly the same section of water is not sampled on each occasion.

On the River Aire, the first, third and fourth sampling occasions (times 1, 3 and 4) were at periods of low dry flow whereas the second sampling occasion (time 2) was immediately following a rain event. The data showed that on all occasions ECL responses increased from low levels at reference Station 1 (<10.3 ECL Units on all occasions), the site of good biological quality based on macroinvertebrate surveys (see Table 1.6) to high levels at Stations 6,7 and 8 (>67.5 ECL Units on all occasions), where biological quality was poor (see Table 1.6). Responses at Station 5 (the positive control) were higher than Station 1 but lower than those at the downstream stations 6, 7 and 8. The ECL values measured at Station 1 and Stations 6, 7 and 8 were typical of those recorded previously at 'clean' and impacted stations respectively (Environment Agency, 1997b).

Table 3.2 Results of ANOVA for the ECL bioassay data from the River Aire and Willow Brook and the ranking of responses

Site	Time	Variance ratio and level of probability				Ranking of responses at stations (lowest to highest)
		Time	Sample	Station	Time vs Station	
River Aire	1-4	23.9 p<0.001	0.4 NS	276.4 p<0.001	4.81 p<0.001	1,2,4,3,5,6,7,8
	1	-	-	660.1 p<0.001	-	1,2,4,3,5,7,6,8
	2	-	-	305.6 p<0.001	-	1,2,4,3,5,6,7,8,
	3	-	-	320.5 p<0.001	-	1,2,4,3,5,7,6,8
	4	-	-	218 p<0.001	-	1,2,4,3,5,7,6,8
Willow Brook	1-3	3.46 p<0.05	1.2 NS	75.9 p<0.001	5.89 p<0.001	4,1=7,3,8,6,5,2
	1	-	-	82.7 p<0.001	-	4,3,7,1,8,6,5,2
	2	-	-	830.4 p<0.001	-	1,3,4,7,8,6,5,2
	3	-	-	59.7 p<0.001	-	7,4,8,1,3,6,2,5

On the Willow Brook no obvious differences in flow between stations were noted on any sampling occasion. In addition, flow conditions appeared to be similar on each sampling occasion. However, this does not preclude the influence of weather events and flow regime in the period leading up to the sampling occasions. The data showed that on all occasions ECL responses were low at Stations 1, 3, 4, 7 and 8 (<9.2 ECL Units in all instances) and increased through Station 6 (14.3-22.7 ECL Units) to highest levels at Stations 2 and 5 (>23.4-44.6 ECL Units on all occasions). Stations 1, 2 and 7 were of the poorest biological quality, while reference Station 8 was of good biological quality (see Table 1.6).

The 3-way ANOVA also demonstrated that for both the River Aire and Willow Brook the differences between samples taken from the same stations were not significant (that is no sample effect). The lack of spatial variability in this test suggests that the five spot water samples were homogenous (see Table 3.2).

Microtox - acute bioassay

Figures 3.3 and 3.4 show the results of Microtox - acute bioassays performed on four occasions for the River Aire and three occasions for the Willow Brook. Water samples were collected at all eight stations at each site on each occasion.

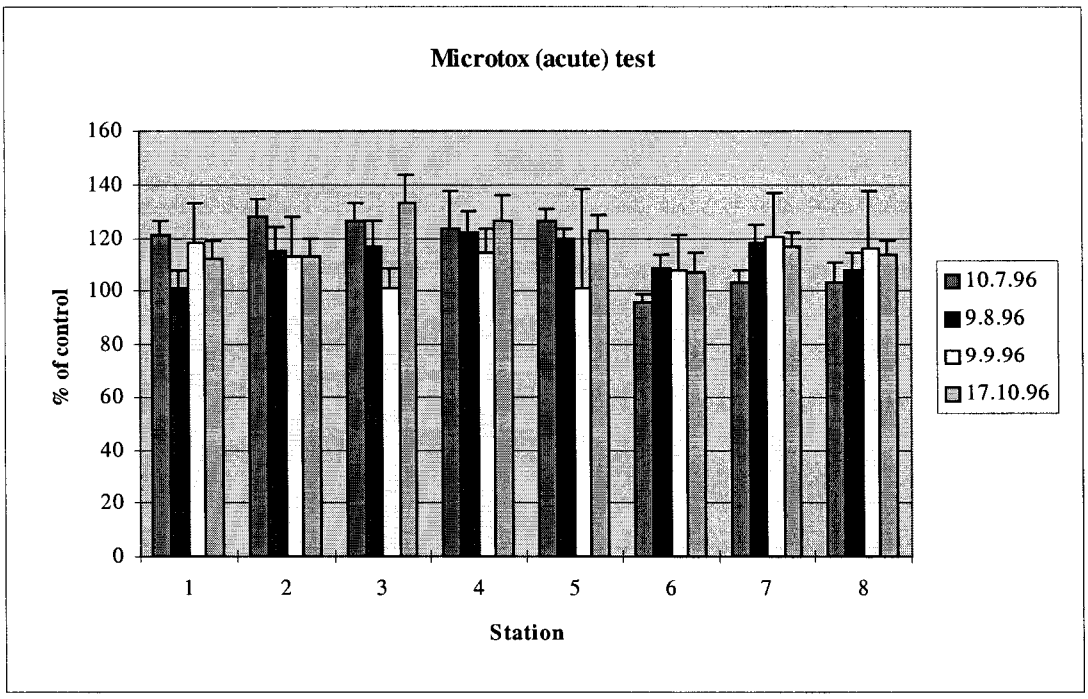

Figure 3.3 Mean Microtox - acute responses on four sampling occasions for the River Aire (error bars represent one standard deviation of station samples from the mean)

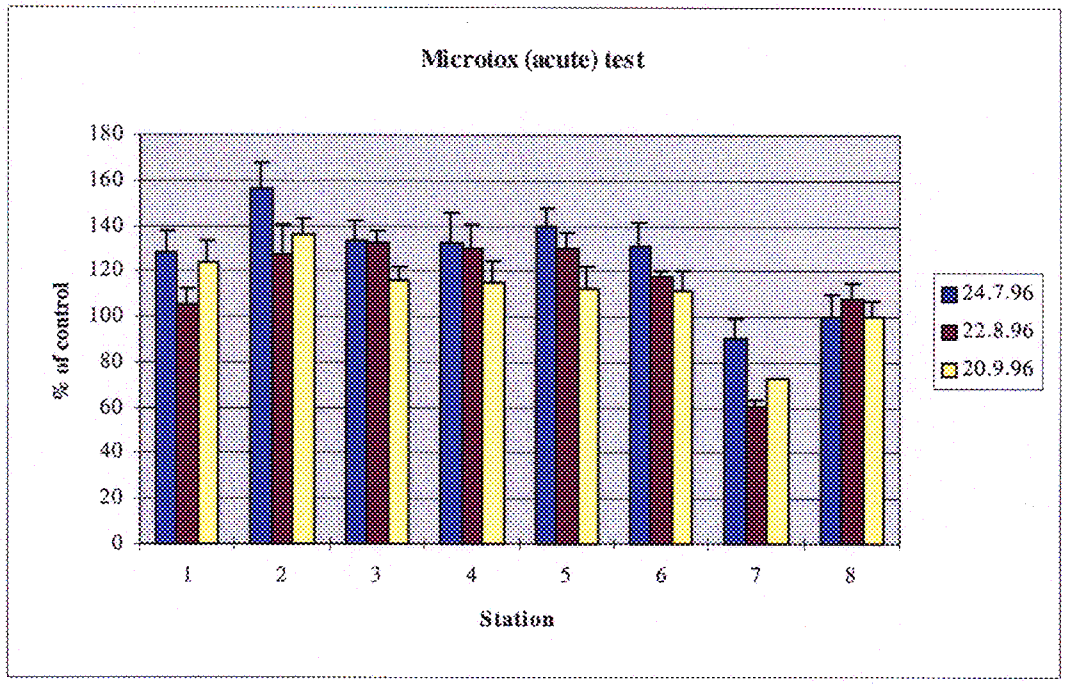

Figure 3.4 Mean Microtox - acute responses three sampling occasions for the Willow Brook (error bars represent one standard deviation of station samples from the mean)

Table 3.3 shows the results of the three and two way ANOVA and the ranking of responses between stations. ANOVA demonstrated there were statistically significant differences between stations at both the River Aire and the Willow Brook. This was the case when both sampling occasions were considered together and separately (see Table 3.3). However, as demonstrated by the station versus time interaction, the differences between stations for both rivers were different for the different sampling occasions.

The 3-way ANOVA showed the differences in mean Microtox - acute responses with sampling date to be statistically significant for both rivers (River Aire: variance ratio = 3.31, $p<0.05$, df=3, Willow Brook: variance ratio = 38.1, $p<0.001$, df=2) respectively. This may be due to flow conditions up to and including the sampling time and patterns in the discharge of contaminants and may include an element of spatial variability in the temporal variability.

On the River Aire, the first, third and fourth sampling occasions (times 1, 3 and 4) were at periods of low flow whereas the second sampling occasion (time 2) was immediately following a rain event. On the Willow Brook no obvious differences in flow were noted at the time of sampling. However, this does not preclude the influence of weather events and flow regime in the period leading up to the sampling occasion.

Table 3.3 Results of ANOVA for the Microtox - acute bioassay data from the River Aire and Willow Brook and the ranking of responses

Site	Time	Variance ratio and level of probability				Ranking of responses at stations (highest to lowest)
		Time	Sample	Station	Time vs Station	
River Aire	1-4	3.31 p<0.05	0.22 NS	5.75 p<0.001	3.79 p<0.001	4,3,5,2,7,1,8,6
	1	-	-	94.8 p<0.001	-	2,5,3,4,1,7,8,6
	2	-	-	37.2 p<0.001	-	4,5,7,3,2,6,8,1
	3	-	-	6.44 p<0.001	-	7,1,8,4,2,6,3,5
	4	-	-	124.9 p<0.001	-	3,4,5,7,8,2,1,6
Willow Brook	1-3	38.1 p<0.001	1.48 NS	90.5 p<0.001	5.43 p<0.001	2,3,5,4,6,1,8,7
	1	-	-	321.1 p<0.001	-	2,5,3,4,6,1,8,7
	2	-	-	284.7 p<0.001	-	3,4,5,2,6,8,1,7
	3	-	-	325.2 p<0.001	-	2,1,3,4,5,6,8,7

The responses recorded at the River Aire and the Willow Brook were in most instances greater than 100% of the control indicating stimulation of light output at these stations (that is hormesis). The hormetic response probably resulted from the presence of low concentrations of trace metals such as copper and zinc in the samples which enhanced the metabolism and light production of the bacterium *V. fischeri* above that measured in the maintenance medium to which the controls are exposed. There was no consistent pattern of response between stations over the four sampling occasions at the River Aire. No marked responses were evident at Station 5 (the positive control). For the Willow Brook toxicity was evident at impacted Station 7 (see Table 1.6) on all occasions while non-impacted Station 8 (see Table 1.6) showed no effects (that is responses similar to the test controls).

The lack of measured toxicity in the water samples was consistent with surveys of the Detroit River, Lake St Clair and the St Lawrence River in Canada which have shown that generally surface water samples taken outside effluent discharge mixing zones for a range of industries have no or limited toxicity (that is <20% light inhibition relative to the control) in the Microtox bioassay (Ribo *et al* 1985, Kaiser *et al* 1988 a,b).

The ANOVA also demonstrated that for both the River Aire and Willow Brook there were no differences between samples taken from the same stations suggesting that the five spot water samples taken at each station were homogenous on all occasions (see Table 3.2).

Lemna minor growth inhibition bioassay

Figures 3.5 and 3.6 show the results of *Lemna minor* growth inhibition bioassays performed on two occasions for the River Aire and Willow Brook. Water samples were collected at all eight stations at each site on each occasion. It was initially planned to collect water samples from the River Aire and Willow Brook on four and three occasions respectively but this was modified on the basis of the results obtained on the first two occasions at each site.

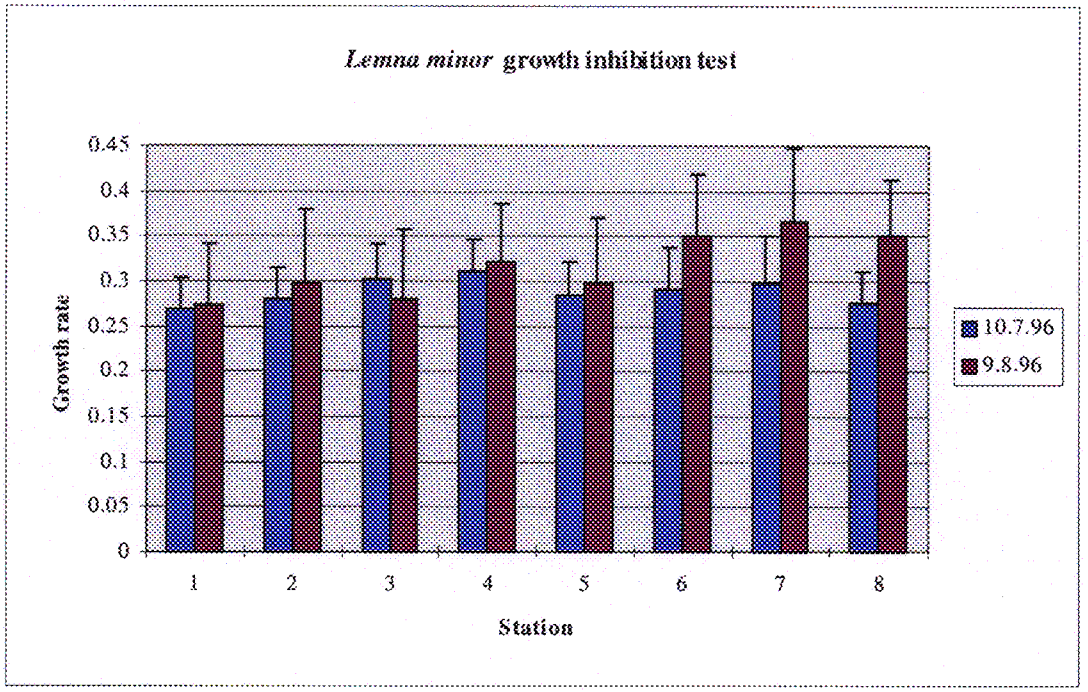

Figure 3.5 Mean *Lemna minor* growth inhibition on two sampling occasions for the River Aire (error bars represent one standard deviation of station samples from the mean)

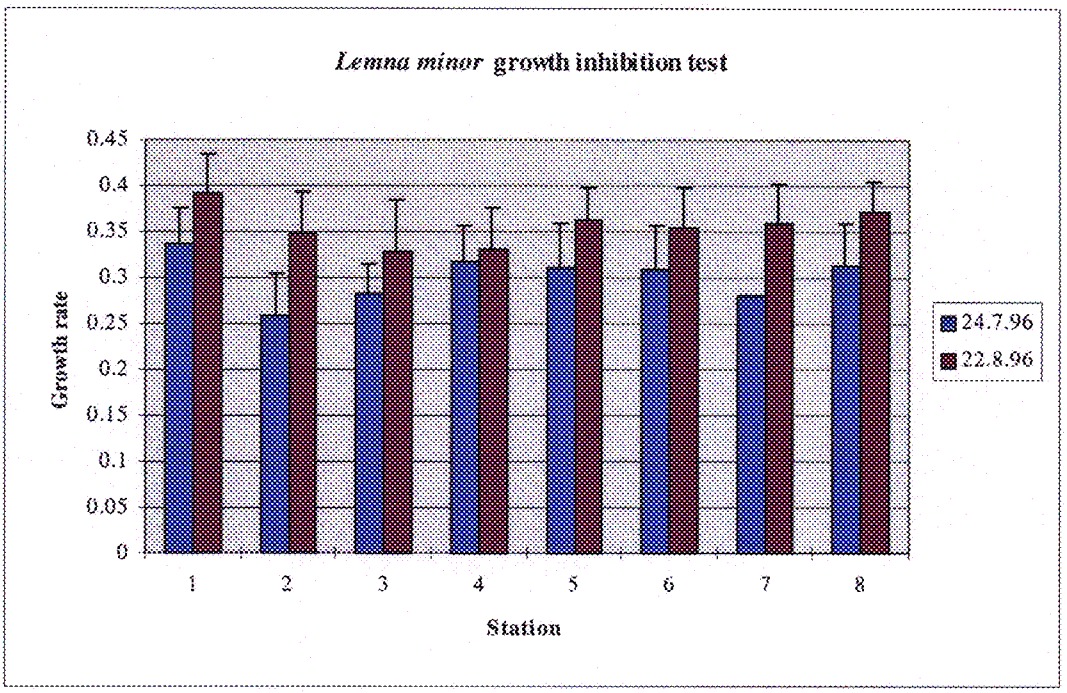

Figure 3.6 Mean *Lemna minor* growth inhibition on two sampling occasions for the Willow Brook (error bars represent one standard deviation of station samples from the mean)

Table 3.4 shows the results of the three and two way ANOVA and the ranking of responses between stations. ANOVA demonstrated that there were statistically significant differences between stations at both the River Aire and the Willow Brook. This was the case when both sampling occasions were considered together and separately (see Table 3.4). However, as demonstrated by the station versus time interaction, the differences between stations for both rivers were different for the two sampling occasions.

The 3-way ANOVA showed the differences in mean *L. minor* growth inhibition responses with sampling date to be statistically significant for both rivers (River Aire: variance ratio = 26.3, $p<0.001$, df=1, Willow Brook: variance ratio = 112.9, $p<0.001$, df=1) respectively. This may be due to flow conditions up to and including the sampling time and patterns in the discharge of contaminants and may include an element of spatial variability in the temporal variability.

Table 3.4 Results of ANOVA for the *Lemna minor* growth inhibition bioassay data from the River Aire and Willow Brook and the ranking of responses

Site	Time	Variance ratio and level of probability				Ranking of responses at stations (highest to lowest)
		Time	Sample	Station	Time vs Station	
River Aire	1 & 2	26.3 p<0.001	0.48 NS	6.35 p<0.001	3.77 p<0.01	7,6,4,8,5,3,2,1
	1	-	-	1.24 NS	-	4,3,7,6,5,2,8,1
	2	-	-	2.38 p<0.05	-	7,6=8,4,2=5,3,1
Willow Brook	1 & 2	112.9 p<0.001	1.38 NS	9.95 p<0.001	2.7 p<0.001	1,8,6,5,4,7,3,2
	1	-	-	4.04 p<0.001	-	1,4,8,5,6,3,7,2
	2	-	-	3.25 p<0.01	-	1,8,5,7,6,2,4,3

At the River Aire, the first sampling occasion (times 1) was at a period of low flow whereas the second sampling occasion (time 2) was immediately following a rain event. No obvious differences in flow were noted at the Willow Brook at the time of sampling. However, this does not preclude the influence of weather events and flow regime in the period leading up to the sampling occasion.

The *L. minor* growth inhibition bioassay measures both nutrient-induced stimulation of growth and toxicant-induced inhibition of growth. There was no consistent pattern of response between stations at either the River Aire or Willow Brook on the two sampling occasions relative to expected contamination gradients.

The ANOVA also demonstrated that for the River Aire and Willow Brook, the differences between samples taken from the same stations were not significant indicating a lack of spatial variability in the five spot water samples (see Table 3.4).

At the River Aire the growth rate of *L. minor* in the laboratory controls (¼ strength Hoaglands medium) was lower than that measured in the receiving water samples from all stations on the first sampling occasions and Stations 4, 6, 7 and 8 on the second sampling occasion (see Table 3.5). These results indicate that the *L. minor* bioassay was generally measuring growth stimulation (eutrophication) in receiving water samples rather than toxicant-indirect growth inhibition. For the Willow Brook the growth in the reference control was only lower than that in receiving water samples from Station 1 on the first sampling occasion and Stations 1, 5, 6, 7 and 8 on the second sampling occasion. These findings indicated that at this site the bioassay was probably measuring both growth inhibition and stimulation.

Table 3.5 Comparison of *Lemna minor* growth in the laboratory reference controls with the receiving water samples from the River Aire and Willow Brook

Site	Time	Growth rate in reference control	Stations showing inhibition related to control	Stations showing stimulation relative to control
River Aire	1	0.232	None	1-8
	2	0.315	1-3, 5	4, 6-8
Willow Brook	1	0.331	2-8	1
	2	0.348	2-4	1, 5-8

Daphnia magna reproduction bioassay

Figures 3.7-3.9 show the results of *D. magna* reproduction bioassay (with juvenile production and lethality as endpoints) performed on four occasions for the River Aire and three occasions for the Willow Brook. Water samples were collected at all eight stations at each site on each occasion.

In the bioassay no lethality was observed with the receiving water samples after two days (that is for a 48 h *D. magna* immobilisation test) and therefore, no specific analysis of this data was appropriate.

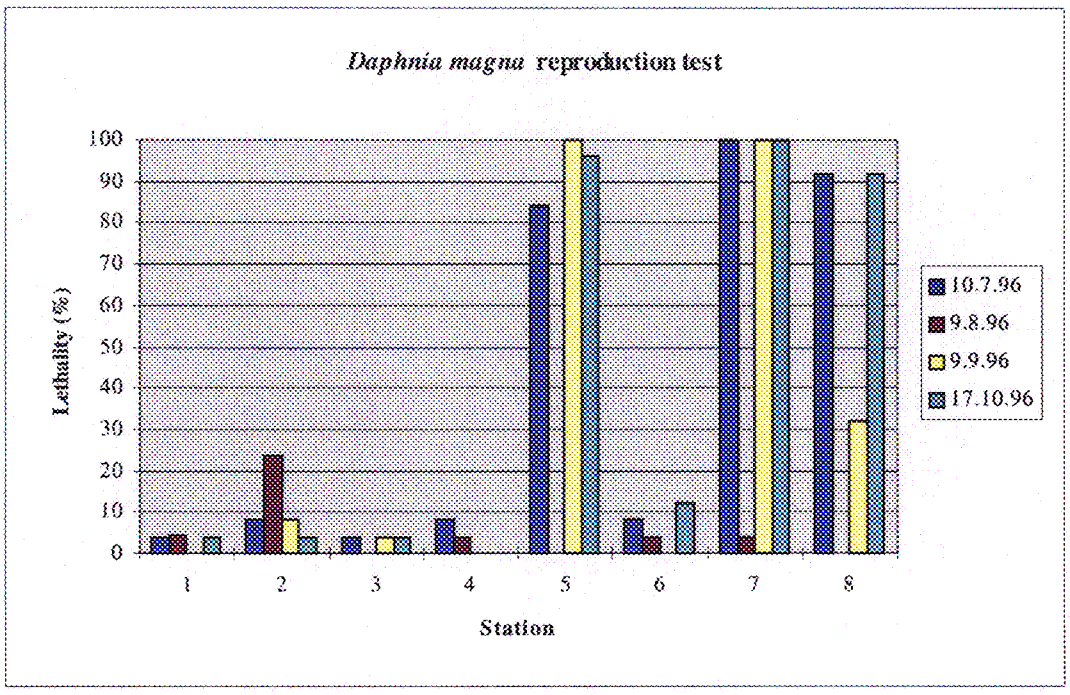

Figure 3.7 Mean *Daphnia magna* juvenile lethality on four sampling occasions for the River Aire (* indicates no lethality measured)

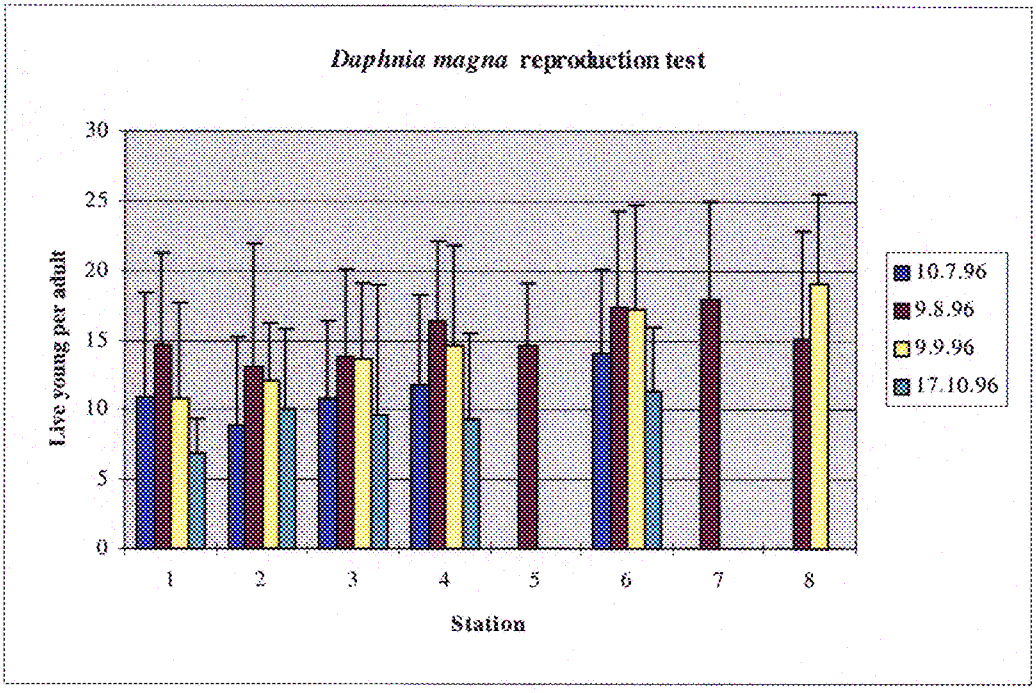

Figure 3.8 **Mean *Daphnia magna* juvenile production on four sampling occasions for the River Aire (error bars represent one standard deviation of station samples from the mean, no production occurred where there was complete lethality)**

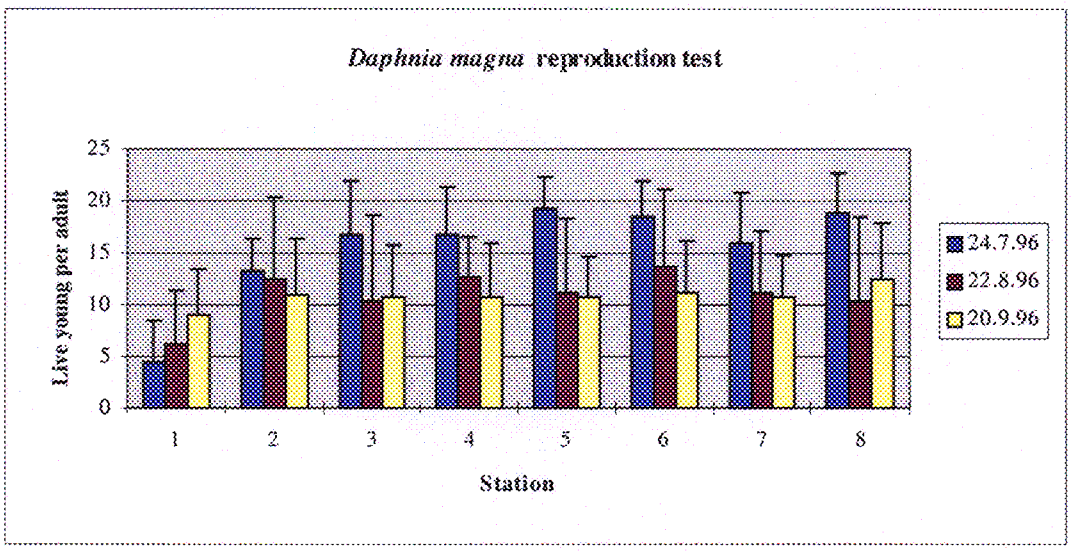

Figure 3.9 **Mean *Daphnia magna* juvenile production on three sampling occasions for the Willow Brook (error bars represent one standard deviation of station samples from the mean)**

Table 3.6 shows the results of the three and two way ANOVA and the ranking of responses between stations. ANOVA demonstrated that there were statistically significant differences between stations at both the River Aire and the Willow Brook. This was the case for both *D. magna* juvenile production and lethality at the River Aire and for *D. magna* juvenile production at the Willow Brook site when the sampling occasions were considered together and separately (see Table 3.6).

Table 3.6 Results of ANOVA for the *Daphnia magna* reproduction bioassay data from the River Aire and Willow Brook and the ranking of responses

Site	Time	Variance ratio and level of probability				Ranking of responses at stations (lowest to highest for lethality, reverse for juvenile production)
		Time	Sample	Station	Time vs Station	
River Aire Lethality	1-4	48.3 p<0.001	1.29 NS	97.1 p<0.001	16.4 p<0.001	3=4,1,6,2,8,5,7
	1	-	-	55.8 p<0.001	-	1=3,2=4,6,5,8,7
	2	-	-	3.86 p<0.01	-	3=5=8,4=6=7,1,2
	3	-	-	58.1 p<0.001	-	1=4=6,3,2,8,5,7
	4	-		139.1 p<0.001	-	4,1=2=3,6,8,5,7
Juvenile production	1-4	90.3 p<0.001	1.98 NS	39.1 p<0.001	9.44 p<0.001	6,4,3,1,2,8,5=7
	1	-	-	1.8 NS	-	8,4,1,3,2
	2	-	-	1.26 NS	-	7,6,4,8,1,5,3,2
	3	-	-	5.11 p<0.001	-	6,4,8,1,3,2
	4	-	-	2.14 NS	-	6,2,3,4,1
Willow Brook Lethality	1-3	3.07 NS	1.5 NS	0.78 NS	0.78 NS	1=2=3=4=5=6=7=8
Juvenile production	1-3	19.0 p<0.001	0.34 NS	7.62 p<0.001	2.28 p<0.001	6,8,5,3,4,7,2,1
	1	-	-	35.9 p<0.001	-	5,8,6,3,4,7,2,1
	2	-	-	4.07 p<0.001	-	6,4,2,5=7,3=8,1
	3	-	-	0.91 NS	-	8,6,2,3=4=5,7,1

The 3-way ANOVA showed these differences in mean *D. magna* responses (juvenile production and lethality) with sampling date to be statistically significant for juvenile production at both rivers (River Aire: variance ratio = 90.3, $p<0.001$, df=3, Willow Brook: variance ratio = 19.0, $p<0.001$, df=2), while sampling date was only significant for lethality at the River Aire site (variance ratio = 48.3, $p<0.001$). This may be due to flow conditions up to and including the sampling time and patterns in the discharge of contaminants and may include an element of spatial variability in the temporal variability.

On the River Aire, the first, third and fourth sampling occasions (times 1, 3 and 4) were at periods of low flow whereas the second sampling occasion (time 2) was immediately following a rain event. The data showed high levels of *D. magna* lethality (generally >84%) at Stations 5, 7 and 8 on the first, third and fourth sampling occasions compared to low levels of lethality (<10%) at all other stations. There was limited (<25%) mortality at all stations on the second sampling occasion. The high levels of lethality recorded at Stations 5, 7 and 8 on the first, third and fourth sampling occasion meant that no or few adults survived to produce juveniles. Therefore, these stations were excluded from the ANOVA of the juvenile production data. As a result juvenile production was compared at Stations 1, 2, 3, 4 and 6 on the first sampling occasion, all stations on the second sampling occasion, Stations 1, 2, 3, 4, 6 and 8 on the third sampling occasion and Stations 1, 2, 3, 4 and 6 on the fourth sampling occasion. No consistent pattern was evident in the number of live juveniles produced at the first brood by surviving adults on the different sampling occasions.

On the Willow Brook there were no obvious differences in flow noted at the time of sampling. However, this does not preclude the influence of weather events and flow regime in the period leading up to the sampling occasion. Lethality of *D. magna* was low (<8%) at all stations on all occasions. The number of live juveniles produced at the first brood by surviving adults was always lower at Station 1 than the other stations.

The ANOVA also demonstrated that for the River Aire and Willow Brook, the differences between samples taken from the same stations were not significant indicating a lack of spatial variability in the five spot water samples (see Table 3.6).

At certain stations on the River Aire the lethality measured at the end of the bioassay (10 days) was evident after six days while at others there was a marked increase in lethality from Days 6 to 10 (see Table 3.7). The increased lethality of adult *Daphnia* from Days 6-10 apparently indicated a strategy of organisms producing a brood of juveniles before succumbing to the toxicants present in the samples.

At both the River Aire and Willow Brook sites the mean number of juveniles per surviving adult produced in the receiving water samples from most stations were higher than the mean number of 8.8 produced in the reference controls over the testing period (see Table 3.8). This finding indicates that at stations where adult lethality was low the receiving water samples resulted in a stimulatory (hormetic) effect. This stimulation of juvenile production may be due to increased food availability in the environmental samples above the optimal levels in the reference controls. The only exceptions were water samples from Station 1 on Willow Brook, which showed decreases in juvenile production on the first two sampling occasions.

Table 3.7 Lethality of *Daphnia magna* in River Aire receiving water samples after different exposure periods

Site	Time	Exposure period (d)	Lethality (%) at given station 1	2	3	4	5	6	7	8
River Aire	1	6	0	4	4	4	72	4	100	12
		10	4	8	4	8	84	8	100	92
	2	6	4	4	0	4	0	4	4	0
		10	4	24	0	4	0	4	4	0
	3	6	0	8	4	9	100	0	12	30
		10	0	8	4	0	100	0	100	30
	4	6	4	0	0	0	0	4	16	30
		10	4	4	4	0	92	12	100	92

Table 3.8 Summary of juvenile production in the *D. magna* reproduction tests at the different River Aire and Willow Brook stations relative to that in groundwater controls

Site	Time	Mean juvenile production in groundwater	Mean juvenile production at each station as % of that in groundwater control 1	2	3	4	5	6	7	8
River Aire	1	9.7	112	92	110	122	-	144	-	-
	2	10.3	143	127	134	158	141	168	174	147
	3	8.9	122	135	153	164	-	193	-	215
	4	6.3	110	146	152	149	-	181	-	-
Willow Brook	1	11.0	40	120	153	152	176	167	146	171
	2	8.0	76	156	130	158	140	170	140	130
	3	8.9	114	123	121	121	121	125	120	140

3.1.2 Freshwater toxicity bioassays - sediments

Chironomus riparius lethality bioassay

Figures 3.10 and 3.11 show the results of *Chironomus* lethality bioassays performed on two occasions for the River Aire and Willow Brook respectively.

For the River Aire no sediment samples were taken for Stations 6 and 7 on either sampling occasion. Station 6 was situated in a stretch of river with coarse substrate and no depositional area could be found. Station 7 was situated in an industrial complex and no access to the sediment was possible. For the Willow Brook, no sediment samples were taken for Station 6 which again was a fast flowing stretch of river with no depositional area.

Table 3.9 shows the results of the three and two way ANOVA and the ranking of responses between stations. ANOVA demonstrated that there were statistically significant differences between stations at both the River Aire and the Willow Brook. This was the case when both sampling occasions were considered together and separately (see Table 3.9). However, as

demonstrated by the station versus time interaction, the differences between stations for both rivers were different for the two sampling occasions.

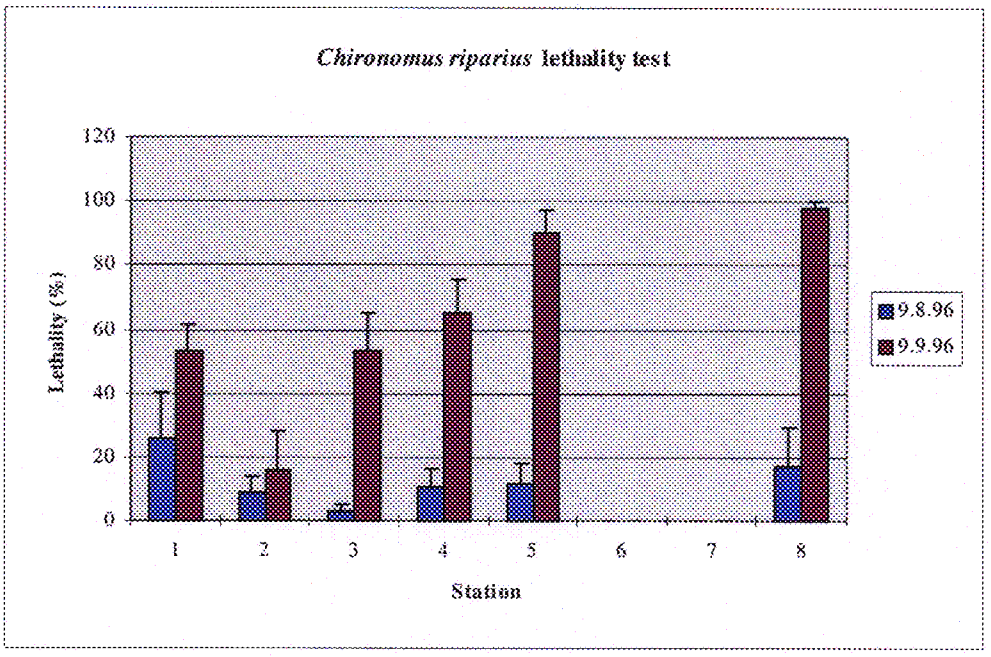

Figure 3.10 Mean lethality (10 d) of *Chironomus riparius* on two sampling occasions for the River Aire (error bars represent one standard deviation of station samples from the mean, NS = no samples taken)

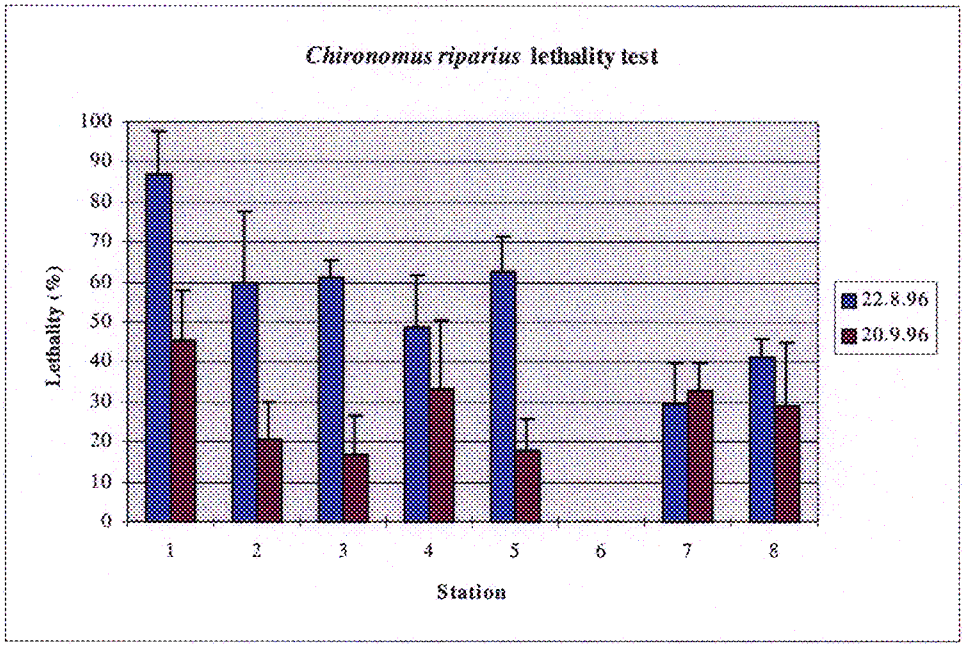

Figure 3.11 Mean lethality (10 d) of *Chironomus riparius* on two sampling occasions for the Willow Brook (error bars represent one standard deviation of station samples from the mean, NS = no samples taken)

Table 3.9 Results of ANOVA for the *Chironomus riparius* lethality bioassay data from the River Aire and Willow Brook and ranking of responses

Site	Time	Variance ratio and level of probability				Ranking of responses at stations (lowest to highest)
		Time	Sample	Station	Time vs Station	
River Aire	2 & 3	392.6 p<0.001	1.61 NS	22.3 p<0.001	17.9 p<0.001	2,3,4,1,5,8
	2	-	-	4.16 p<0.01	-	3,2,4,5,8,1
	3	-	-	63.9 p<0.001	-	2,1=3,4,5,8
Willow Brook	2 & 3	100.1 p<0.001	0.1 NS	8.82 p<0.001	6.06 p<0.001	7,8,3,5,2,4,1
	2	-	-	21.1 p<0.001	-	7,8,4,2,3,5,1
	3	-	-	7.95 p<0.001	-	3,5,2,8,7,4,1

Mean lethality measured in both the River Aire and Willow Brook samples was different for the two sampling occasions. The 3-way ANOVA showed these differences in sampling date to be statistically significant for both rivers (River Aire: variance ratio =392.6, $p<0.001$, df=1; Willow Brook: variance ratio = 100.1, $p<0.001$, df=1 respectively). This may have been due to flow conditions up to and including the sampling time and changes in the discharge of contaminants. It may include an element of spatial variability within the temporal variability observed as it is not possible to sample the same sediment twice and there may be heterogeneity of sediment type on small spatial scales.

On the River Aire, the first sampling occasion was at a period of higher flow immediately following a rain event. This may have been responsible for a reduction in sediment toxicity due to dilution of contaminants and scouring of contaminated sediment. At all stations, lethality was $\leq$23%, which is similar to the level of 20% commonly accepted to be equivalent to control toxicity and interpreted as 'no response'. At Station 1 the mean toxicity was influenced by high lethality in one replicate of sample 2 which when excluded resulted in mean lethality of 19%.

The second sampling time was at low flow following a period of dry weather. At this time, significant toxicity, that is above 20%, was observed for all sites with the exception of Station 2. On this occasion, with the exception of Station 1, lethality increased with distance downstream. The highest toxicity (>80% lethality) was recorded for Station 8 which was of poor quality (see Table 1.6). More than 80% lethality was also measured at the positive control Station 5 (see Table 1.6) which was immediately downstream of a sewage discharge and the sediments were observed to be anoxic and sulphurous. The high toxicity of sediments at Station 1 on the second sampling occasion was unexpected given the low toxicity at downstream station 2 and the absence of toxic effects on the first occasion. Station 1 was surrounded by farmland and it is possible that run-off of hydrophobic contaminants such as

pesticides had occurred during a rain event, or that aqueous discharge from the farm had entered the system.

On the Willow Brook no obvious differences in flow were observed on the two sampling occasions. However, this does not preclude the influence of weather events and flow regime in the period leading up to both sampling times. In general, greater toxicity was observed on the first occasion. On the second occasion <50% lethality was seen at all stations. Station 1 was classed as being the most toxic on both occasions which is consistent with the historical biological data (see Table 1.6). On the first occasion, Station 8 demonstrated the least toxicity which is consistent with the historical biological survey data. However, even at this site >20% lethality was observed suggesting that some sediment toxicity may exist.

The ANOVA demonstrated that for both the Aire and the Willow Brook, differences between spot samples taken from within the same station were not significant in influencing the patterns of toxicity (see Table 3.9). A comparison of spot samples from each station gives an indication of the spatial variability in test response. Given the large spatial variability in sediment type on a small horizontal and vertical scale, variability in sediment toxicity between samples is often expected. However, the lack of spatial variability in this bioassay suggests that either the sediment type did not vary between the five spot samples or that sediment type did not significantly influence the bioassay response. Tables 3.10 and 3.11 show summary information for the physico-chemical determinands measured in spot samples for the second sampling occasion on the River Aire and both the sampling occasions on the Willow Brook. The data are summarised as means of the five samples, and the standard deviation indicates the variation at each station. The particle size data is summarised as the percentage in the <63μm fraction which is equivalent to the silt/clay fraction. This fraction is thought to contain the biologically active material and influence contaminant partitioning and bioavailability (Ingersoll 1996).

Table 3.10 Summary of physico-chemical determinands measured in River Aire sediments on the second sampling occasion

Time-station	TOC (%)		Particle size (<63 μm) (%)		Water content (%)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
1-1 to 1-8			Samples lost in transit			
2-1	3.3	0.8	44.0	4.9	51.5	3.3
2-2	2.3	0.2	42.8	6.6	44.0	5.6
2-3	1.4	0.6	25.9	3.0	34.4	4.6
2-4	2.0	1.8	18.7	13.1	38.7	17.8
2-5	2.3	1.1	9.7	1.5	34.6	3.2
2-6			No sample			
2-7			No sample			
2-8	2.7	1.5	15.2	4.2	30.6	3.0

The samples from the first sampling occasion on the River Aire were lost by couriers in transit to the laboratory for particle size analysis.

Table 3.11 Summary of physico-chemical determinands measured in Willow Brook sediments on both sampling occasions

Time-	TOC (%)		Particle size (<63 μm) (%)		Water content (%)	
station	Mean	S.D.	Mean	S.D.	Mean	S.D.
1-1	1.7	1.3	28.2	14.9	-	-
1-2	2.6	1.5	63.1	7.1	-	-
1-3	1.5	0.7	53.7	9.7	-	-
1-4	6.0	1.9	59.0	14.1	-	-
1-5	2.6	2.3	23.1	20.2	-	-
1-6			No sample			
1-7	5.8	0.3	15.2	0.9	-	-
1-8	5.7	0.5	57.3	25.3	-	-
2-1	1.6	0.5	18.2	1.5	44.5	6.9
2-2	4.1	1.3	62.5	4.5	59.2	8.8
2-3	2.0	0.8	58.2	12.6	55.2	9.8
2-4	5.5	1.9	55.0	15.5	62.9	5.1
2-5	1.7	0.6	23.2	7.8	41.8	15.0
2-6			No sample			
2-7	6.8	0.5	25.5	4.0	60.0	3.9
2-8	4.9	0.3	31.2	7.8	69.1	14.6

The standard deviation of each physico-chemical determinand at each station indicate that there was generally a low level of variability (<25%) in sediment type between spot samples. However, this variation was probably not sufficient to cause significant differences in the toxic response between the samples.

Although spatial variability did not seem to affect bioassay response within stations, variation in sediment type may have been influencing the significant differences in bioassay response observed between stations and, therefore, contributing to the patterns of toxicity. The influence on test response of levels of toxicants and the natural stress of sediment type is difficult to differentiate. Several approaches have been suggested for normalising toxic responses on the basis of sediment factors (for example DeWitt *et al* 1988, Ankley *et al* 1994). One approach is to construct a regression model for each sediment factor against the response of interest. The regression line is then overlaid onto the response measured in various test sediments. This is the most rigorous method of sediment type normalisation. However, to construct the model, toxicity tests and physico-chemical measurements must be performed on a large number of uncontaminated sediments. An alternative is to conduct parallel testing with reference sediments which are similar in physico-chemical character but uncontaminated with toxicants. Any response measured in the reference sediment can, therefore, be attributed to sediment type. This approach is less rigorous than a regression approach because it assumes that the sediment factors influencing toxic response are known and can be matched with the test sediments. It also becomes logistically difficult when testing a large number of sediments (as in this study). An option would have been to include two reference sediments which covered the ranges in physico-chemical determinands measured in the test sediments. However, because of the testing requirements of this study reference sediments such as these

were not included due to logistical constraints. In order to investigate the influence of sediment type here a third approach was used, similar to that of the regression approach but not as statistically rigorous. The approach involved comparing the ranges of physico-chemical determinands measured in the test sediments with those characteristics which have been shown in previous tests with uncontaminated sediments to support *Chironomus* survival for periods of up to 28 days. Only those sediments were considered which gave rise to less than 20% lethality over a ten day period in which the test animals were unfed, or longer term (28 day) tests in which the animals were fed. Such ranges of TOC, silt/clay fraction and water content are shown in Table 3.12.

A comparison of these ranges with the mean values for the test sediments in Tables 3.10 and 3.11 shows that in general, particle size and TOC measured in the test sediments fell within ranges which have previously supported *Chironomus* survival. Insufficient data exist for a similar comparison with water content. The one exception in this comparison is the TOC content of the sediments at Stations 4 and 7 on the Willow Brook. It is possible that TOC may have affected the toxic response at these stations. With the exception of these stations, it is likely that the pattern in toxic response between stations can be largely attributed to either toxicant concentration or to other unidentified factors which were not measured in this study. Table 3.13 shows the results of analyses of correlations between the 10 day lethality data and sediment physico-chemical parameters. At the River Aire no statistically significant correlations were evident between lethality and TOC or water content on the second sampling occasion while there was a significant correlation ($r = -0.681$, $p<0.001$, $n = 28$) between lethality and the 63 μm particle size fraction. The negative correlation indicated that lethality increased as the % of <63 μm particles in the sediment decreased.

Table 3.12 Ranges of physico-chemical factors which support 10 day *Chironomus* survival

Sediment factor	Range (%)	Sediment type	Test duration	Reference
TOC	0.2 - 3.0	artificial	10 d	Fleming and Nixon (1996)
	0.5 - 3.0	artificial	28 d	"
	1.2	natural	28 d	"
	5.4	natural	10 d	Fleming *et al* (1994)
<63 µm fraction	25 - 75	artificial	10 d	Fleming and Nixon (1996)
	20 - 50	artificial	28 d	"
	52	natural	28 d	"
	5.2	natural	10 d	(Holmes *et al* (1995)
	2.2	natural	10 d	(Fleming *et al* (1994)
Water content	33	artificial	10 d	Fleming and Nixon (1996)
	38 - 42	artificial	28 d	"
	28 - 44	artificial	28 d	"

At the Willow Brook the only statistically significant correlation established was between lethality and TOC data ($r = -0.647$, $p<0.001$ $n = 33$). The negative correlation indicated that

lethality increased as the TOC content decreased. Whether this was due to the intercorrelation between toxicant bioavailability and TOC or to the natural stress of decreasing nutritive quality of the sediment is not possible to determine without further analysis of toxicant concentrations and a more thorough investigation of the influence of sediment factors. However, the comparison with previous sediment type data would suggest that the sediment types encountered in this study should have supported the survival of *Chironomus* if they were uncontaminated.

Table 3.13 Correlation coefficients (and significance) between 10 day lethality data and sediment physico-chemical parameters

Site	Time	Correlation coefficients (and significance) between lethality data and physico-chemical parameters		
		TOC	Particle size (<63 μm)	Water content
River Aire	2	ND	ND	ND
	3	0.139 (NS)	-0.681 (p<0.001)	-0.309 (NS)
Willow Brook	2	-0.647 (p<0.001)	ND	-0.194 (NS)
	3	0.242 (NS)	-0.06 (NS)	-0.254 (NS)

ND = no data, NS = not significant

Chironomus riparius growth/lethality bioassay

Figures 3.12 and 3.13 show the results of *Chironomus riparius* growth/lethality performed on one occasion for the River Aire.

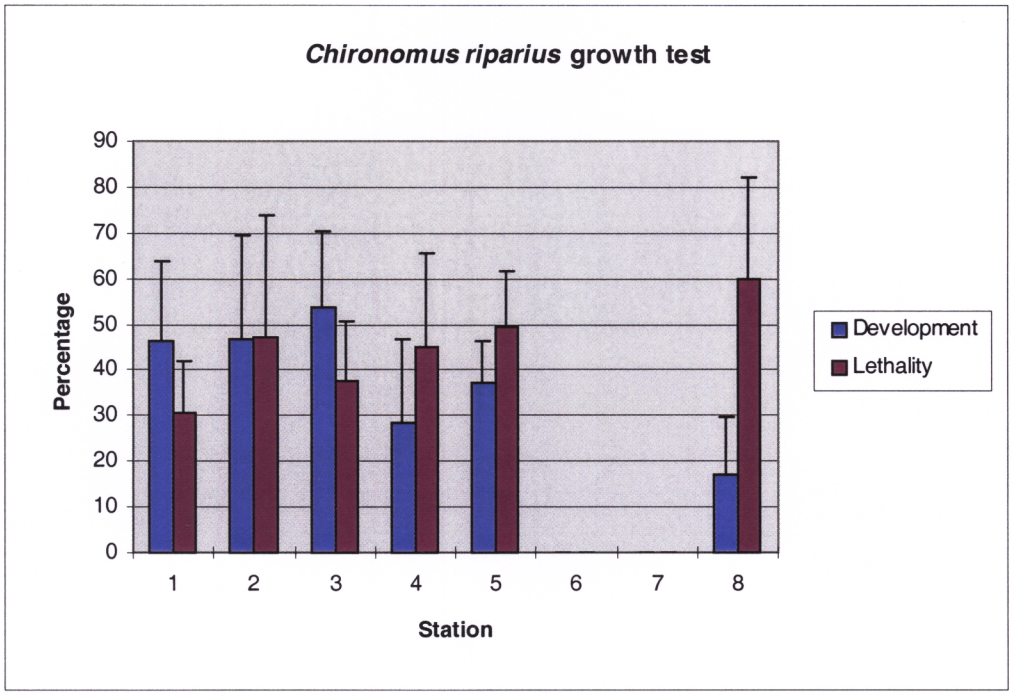

Figure 3.12 Mean lethality (28 d) and development of *Chironomus riparius* on one sampling occasion for the River Aire (error bars represent one standard deviation of station samples from the mean, NS = no samples taken)

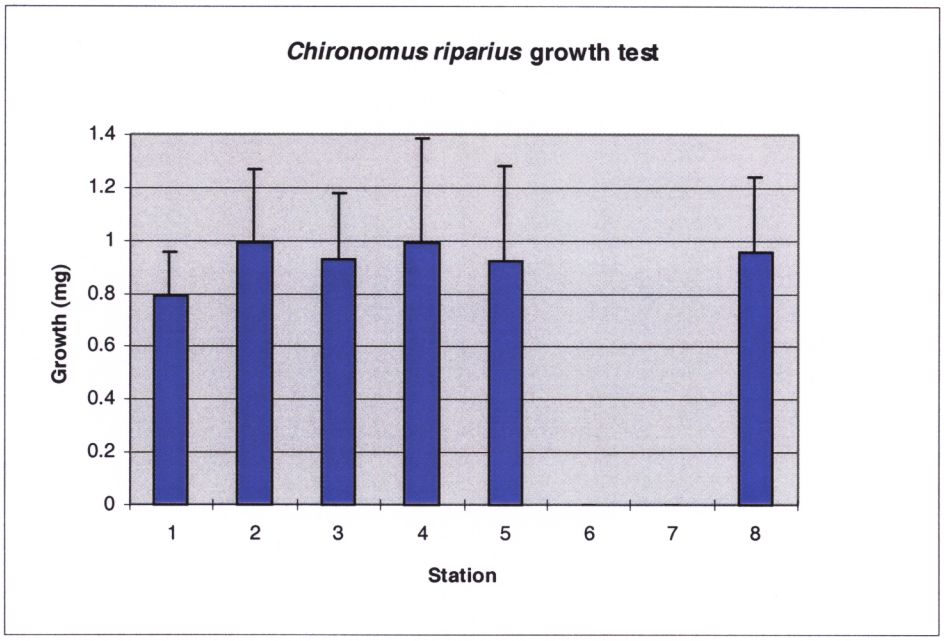

Figure 3.13 Mean growth (as dry weight) of *Chironomus riparius* on one sampling occasion for the River Aire (error bars represent one standard deviation of station samples from the mean, NS = no samples taken)

The results of the second bioassay on the River Aire were not included in the statistical analysis and have not been presented here because it was considered to have failed with high variability being seen between bioassay replicates, particularly those from the reference location. This may have been due to the quality of the test organisms or the influence of test operators and indicates the need for further standardisation of the procedure.

Table 3.14 shows the results of the three and two way ANOVA and the ranking of responses for *Chironomus* lethality, development and growth after 28 days exposure to sediments.

Table 3.14 Results of ANOVA for the *Chironomus riparius* lethality development and growth data from the River Aire and the ranking of responses

Effect	Variance ratio and level of probability					Ranking of responses at stations (lowest to highest for lethality and development, reverse for growth)
	Time	Time	Sample	Station	Time vs Station	
Lethality	2	-	6.54 p<0.001	8.97 p<0.001	-	1,3,4,2,5,8
	3	Test failed				
Development	2	-	13.61 p<0.001	13.31 p<0.001	-	8,4,5,1,2,3
	3	Test failed				
Growth	2	-	7.16 p<0.001	2.31 NS	-	4,2,8,3,5,1
	3	Test failed				

Two-way ANOVA demonstrated that differences in lethality between stations were significant in determining the spread of toxic response observed (variance ratio = 8.97, $p<0.001$, df=5). As expected, development (that is number of larvae developing to fourth instar and above) increased as lethality decreased. The two parameters were significantly negatively correlated ($r= -0.70$, $p<0.001$). Development was lowest at Station 8 where lethality was the greatest. ANOVA demonstrated that there were statistically significant differences in development between stations, with a variance ratio of 13.31 ($p<0.001$, df=5).

For both the lethality and development endpoints in this test, differences between samples, that is spatial variability, were also influential in determining toxic response (lethality variance ratio = 6.54, $p<0.001$, development variance ratio = 13.61, $p<0.001$ respectively) (see Table 3.14). Because differences were observed within stations, the influence of sediment type in determining patterns between stations is likely to be more pronounced in the 28 day bioassay than in the 10 day bioassay. Therefore, the influence of spatial variability on these endpoints should be carefully considered if the tests are used for receiving water monitoring. A regression model, as described earlier, may be one way of overcoming this problem. Lethality over 28 days was significantly correlated with particle size (<63μm, $r=-0.45$, $p<0.05$, n=28) and water content ($r=0.41$, $p<0.05$, n=28). Both factors were negatively correlated which indicates that 28 day lethality, decreases as the water content and silt/clay fraction increased. Development was correlated with the same factors as lethality, that is particle size (<63 μm, $r=0.51$, $p<0.05$, n=28) and water content ($r=0.35$, $p<0.05$, n=28). No significant correlations were evident between TOC and either lethality or development after 28 days exposure. The influence of sediment type versus toxicant concentration cannot be identified further without analysis of toxicant concentrations in the sediments and a thorough investigation of the influence of different sediment factors on control survival.

No obvious patterns of toxicity were observed for the growth endpoint and the ANOVA showed that there was no influence of station on the effects observed. In addition, spatial variability for this endpoint was significant with a variance ratio of 7.16 ($p<0.001$, df=5). Many problems have been observed with growth as a sub-lethal endpoint in the spiked sediment tests performed in previous studies (Fleming and Sims 1997). These studies showed that growth tended to increase as lethality increased since additional resources become available for the survivors. This is supported by the results of this study which demonstrate that 28 day growth is positively correlated with 28-day lethality ($r=0.51$, $p<0.01$). In order to demonstrate an adverse biological effect growth would be expected to decrease with an increase in lethality. This raises questions about the biological significance of growth as measured in these tests systems and further investigation of the factors influencing the response should be investigated further before it could be considered as suitable for receiving water monitoring. Analysis of correlations between the growth endpoints and sediment physico-chemical parameters showed a significant correlation between TOC ($r=-0.502$, $p<0.01$) but no correlations between growth and the <63 μm particle size fraction or water content. The negative correlation indicated that growth increased as TOC decreased although the ranges of both parameters were limited.

3.1.3 Estuarine toxicity bioassays - water column

Enhanced chemiluminescence (ECL) bioassays

Figures 3.14 and 3.15 show the results of ECL bioassays performed on four occasions for the Tees Estuary and three occasions for the Mersey Estuary. Water samples were collected at all eight stations at each site on each occasion.

Table 3.15 shows the results of the three and two way ANOVA and the ranking of responses between stations. ANOVA demonstrated that there were statistically significant differences between stations at both the Tees and Mersey estuaries. This was the case when both sampling occasions were considered together and separately (see Table 3.15). However, as demonstrated by the station versus time interaction, the differences between stations for both estuaries were different for the different sampling occasions.

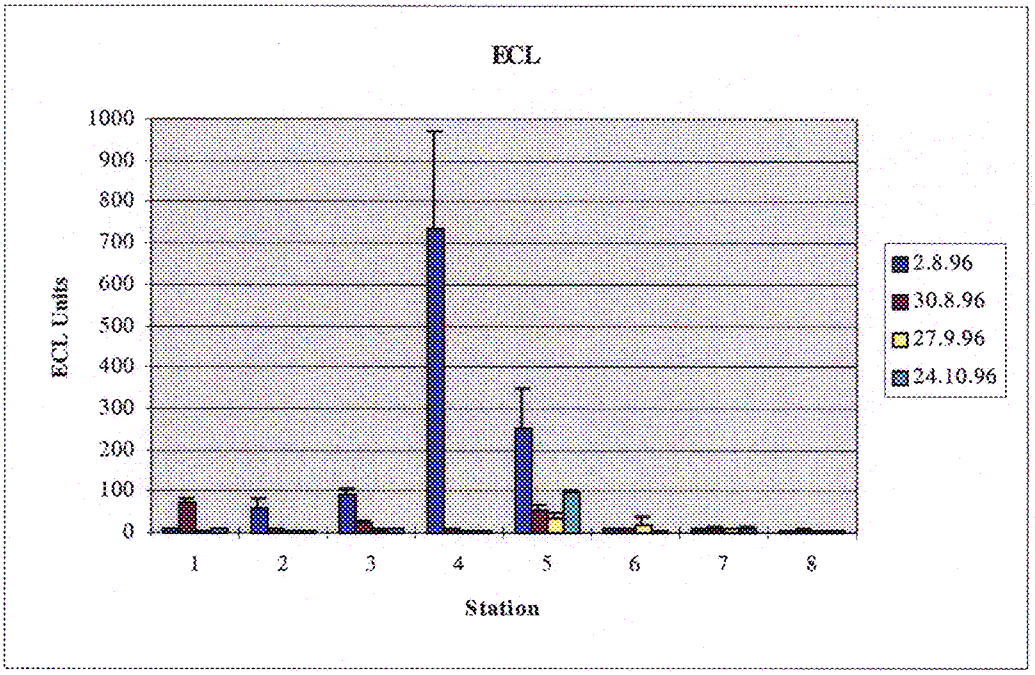

Figure 3.14 Mean ECL responses on four sampling occasions for the Tees Estuary (error bars represent one standard deviation of station samples from the mean)

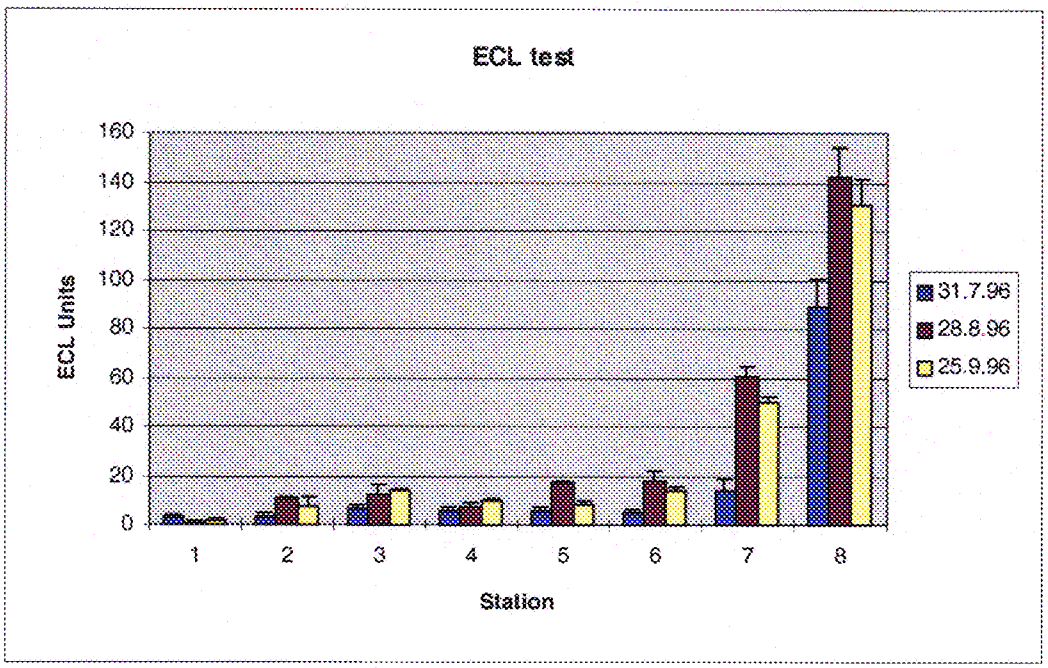

Figure 3.15 Mean ECL responses on three sampling occasions for the Mersey Estuary (error bars represent one standard deviation of station samples from the mean)

The 3-way ANOVA showed the differences in mean ECL responses with sampling date to be statistically significant for both estuaries (Tees Estuary: variance ratio = 261.5, $p<0.001$, df=3, Mersey Estuary: variance ratio = 167.5, $p<0.001$, df=2) respectively. This may be due to changing tidal conditions at the sampling time and patterns in the discharge of contaminants. In addition, an element of spatial variability may be included in the temporal variability as the same section of water was not sampled on each occasion.

The data for the Tees Estuary showed that there were no clear patterns of response between stations, though on all occasions ECL responses were lowest at Stations 7 and 8 (<11.7 ECL Units on all occasions), with the highest levels at Station 5 (> 36.2 ECLOX Units on most occasions). Station 8 represented the location of highest biological quality while Station 5 represented the lowest biological quality location and is regarded as a positive control (see Table 1.6).

On the Mersey Estuary ECL responses were lowest at Station 1 (<3.1 ECL Units in all instances), increased at Stations 2,3,4,5 and 6 (3-18.1 ECL Units) to higher levels at Stations 7 (13.5-60.3 ECL Units) and 8 (>89.3 ECL Units on all occasions).

Table 3.15 Results of ANOVA for the ECL bioassay data from the Tees and Mersey estuaries and the ranking of responses

Site	Time	Variance ratio and level of probability				Ranking of responses at stations (lowest to highest)
		Time	Sample	Station	Time vs Station	
Tees Estuary	1-4	261.5 p<0.001	0.51 NS	146.2 p<0.001	97.2 p<0.001	8,7,6,2,1,3,4,5
	1	-	-	861.6 p<0.001	-	8,7,1,6,2,3,5,4
	2	-	-	420.7 p<0.001	-	8,7,6,4,2,3,5,1
	3	-	-	513.9 p<0.001	-	8,4,2,1,3,7,6,5
	4	-	-	3731 p<0.001	-	8,4,2,6,3,1,7,5
Mersey Estuary	1-3	167.5 p<0.001	0.94 NS	882.9 p<0.001	20.8 p<0.001	1,2,4,5,3,6,7,8
	1	-	-	1154 p<0.001	-	2,1,6,4,5,3,7,8
	2	-	-	3816 p<0.001	-	1,4,2,3,5,6,7,8
	3	-	-	4236 p<0.001	-	1,2,5,4,3,6,7,8

The ANOVA also demonstrated that for both the Tees and Mersey estuaries, the differences between samples taken from the same stations were not significant suggesting that the five spot water samples taken at each station were homogenous (see Table 3.15).

Microtox - acute bioassay

Figures 3.16 and 3.17 show the results of Microtox - acute bioassays performed on four occasions for the Tees Estuary and three occasions for the Mersey Estuary. Water samples were collected at all eight stations at each site on each occasion.

Table 3.16 shows the results of the three and two way ANOVA and the ranking of responses between stations. ANOVA demonstrated that there were statistically significant differences between stations at both the Tees and Mersey estuaries. This was the case when both sampling occasions were considered together and separately (see Table 3.16). However, as demonstrated by the station versus time interaction, the differences between stations for both estuaries were different for the different sampling occasions.

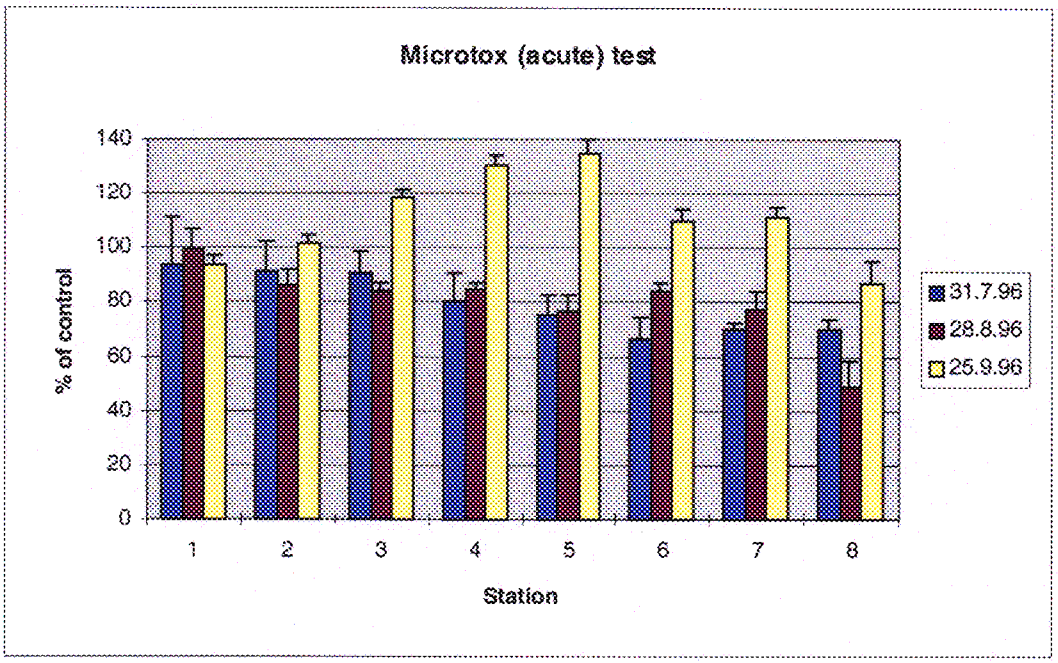

Figure 3.17 Mean Microtox - acute responses three sampling occasions for the Mersey Estuary (error bars represent one standard deviation of station samples from the mean)

The 3-way ANOVA showed the differences in mean Microtox - acute responses with sampling date to be statistically significant for both estuaries (Tees Estuary: variance ratio = 126.8, $p<0.001$, df=3, Mersey Estuary: variance ratio = 274, $p<0.001$, df=2) respectively. This may be due to tidal conditions at the sampling time and patterns in the discharge of contaminants and may include an element of spatial variability in the temporal variability.

On the Tees Estuary the data showed that on all occasions Microtox - acute responses (as % of the light production in the controls) decreased from levels around 100% at Station 1 to lowest levels (11.2-33.8 % of the control) at Station 5 (the positive control). Intermediate effects were measured at Station 8 which was regarded as the reference location.

On the Mersey Estuary the data showed different levels of responses on different sampling occasions. Stations 2-7 showed limited toxicity on the first and second sampling occasions with elevated light production on the third sampling occasion indicating hormesis. Station 1 always showed values similar to the control (93.3-99.2 %) while Station 8 showed toxic effects on all occasions.

The ANOVA also demonstrated that for both the Tees and Mersey estuaries, the differences between samples taken from the same stations were not significant suggesting that the five spot water column samples were homogenous (see Table 3.16).

Oyster embryo-larval development bioassay

Figures 3.18 and 3.19 show the results of the OEL bioassays (Percent Net Responses) performed on four occasions for the Tees Estuary and three occasions for the Mersey Estuary. Water samples were collected at all eight stations at each site on each occasion.

Table 3.17 shows the results of the three and two way ANOVA and the ranking of responses between stations. ANOVA demonstrated that there were statistically significant differences between stations at both the Tees and Mersey estuaries. This was the case when both sampling occasions were considered together and separately (see Table 3.17). However, as demonstrated by the station versus time interaction, the differences between stations for both estuaries were different for the different sampling occasions.

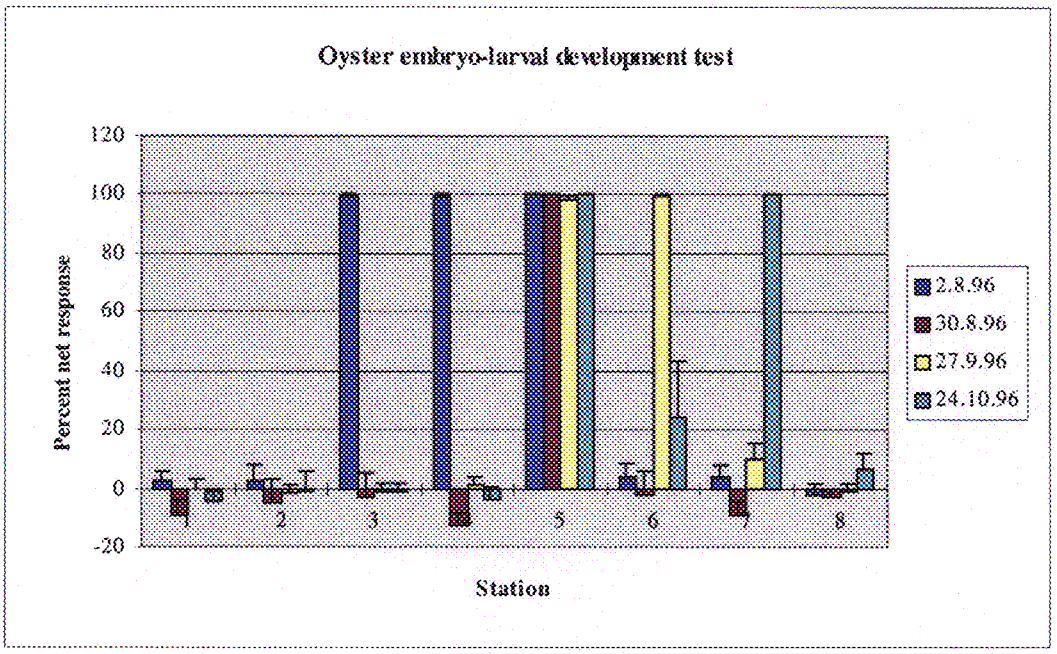

Figure 3.18 Mean OEL responses on four sampling occasions for the Tees Estuary (error bars represent one standard deviation of station samples from the mean)

Table 3.16 Results of ANOVA for the Microtox - acute bioassay data from the Tees and Mersey estuaries and the ranking of responses

Site	Time	Variance ratio and level of probability				Ranking of responses at stations (highest to lowest)
		Time	Sample	Station	Time vs Station	
Tees Estuary	1-4	126.8 p<0.001	1.16 NS	1284 p<0.001	145.6 p<0.001	1,2,4,7,8,3,6,5
	1	-	-	2347 p<0.001	-	1,2,4,8,7,6,3,5
	2	-	-	1705 p<0.001	-	1,2,7,8,4,3,6,5
	3	-	-	8336 p<0.001	-	1,3,2,4,8,7,6,5
	4	-	-	1129 p<0.001	-	2,1,4,6,3,8,7,5
Mersey Estuary	1-3	274 p<0.001	0.82 NS	36.5 p<0.001	21.5 p<0.001	2,4,3,5,1,6,7,8
	1	-	-	115.9 p<0.001	-	1,2,3,4,5,8,7,6
	2	-	-	617.6 p<0.001	-	1,2,4,3=6,7,5,8
	3	-	-	380 p<0.001	-	5,4,3,7,6,2,1,8

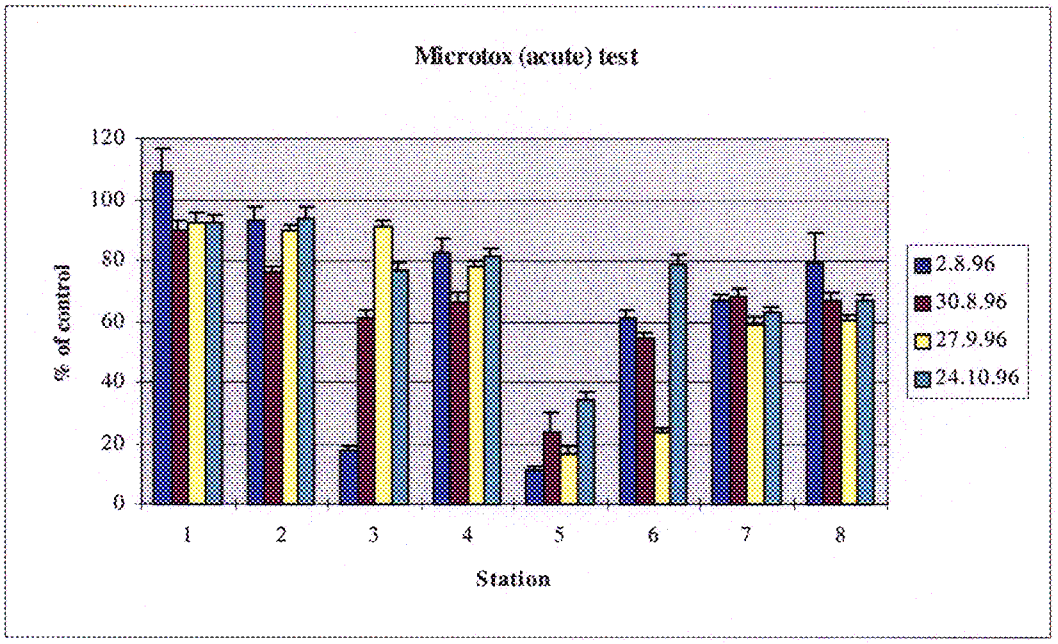

Figure 3.16 Mean Microtox - acute responses on four sampling occasions for the Tees Estuary (error bars represent one standard deviation of station samples from the mean)

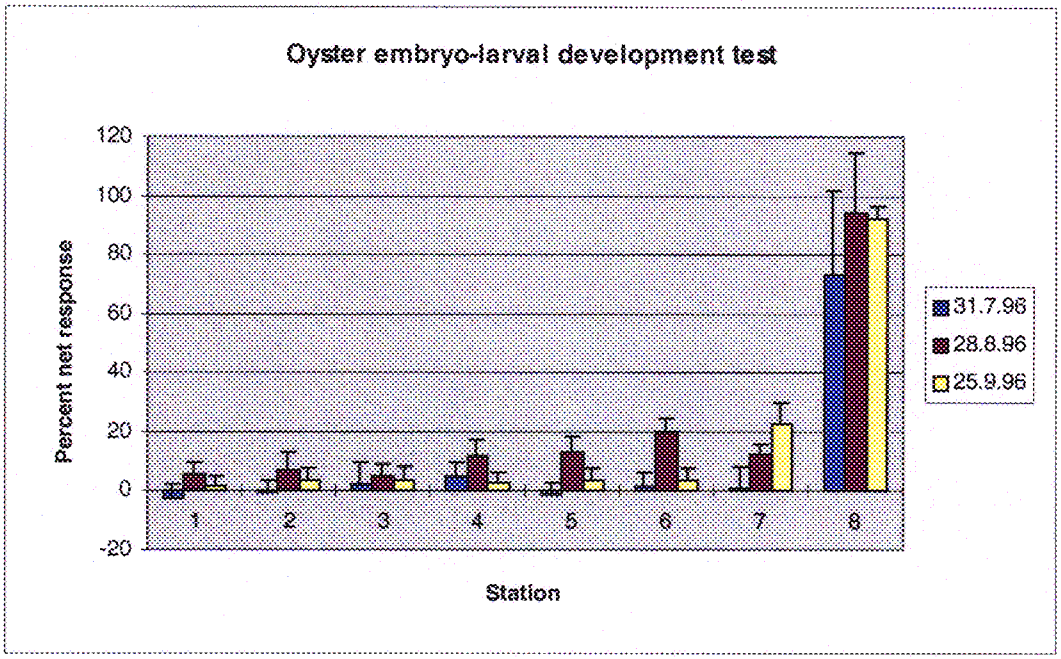

Figure 3.19 Mean OEL responses on three sampling occasions for the Mersey Estuary (error bars represent one standard deviation of station samples from the mean)

The 3-way ANOVA showed the differences in mean OEL PNR responses with sampling date to be statistically significant for both rivers (Tees Estuary: variance ratio = 133.4, $p<0.001$, df=3, Mersey Estuary: variance ratio = 28.8, $p<0.001$, df=2) respectively. This may be due to tidal conditions at the sampling time and patterns in the discharge of contaminants and may include an element of spatial variability in the temporal variability.

On the Tees Estuary, the OEL data showed that on all occasions PNR values were low at Stations 1, 2 and 8 (<6.6), high at Station 5 (>98), the positive control, on all occasions and high on one sampling occasion at Stations 3, 4, 6 and 7. The high values at Stations 3 and 4 on the first sampling occasion were taken slightly after low tide and probably reflected the movement of water from the discharge at positive control Station 5 to these stations. The high values at Stations 6 and 7 on the third and fourth sampling occasions were taken on the falling tide and reflected the movement of water from the discharge at Station 5 towards the open sea.

On the Mersey Estuary OEL PNR values were low at Stations 1-5 (<13.1), higher on one occasion at Stations 6 and 7 (20.2-22.4) and high at Station 8 on all occasions (>73.2).

The ANOVA also demonstrated that for the both the Tees and Mersey estuaries, the differences between samples taken from the same stations were not significant suggesting that the five spot water column samples were homogenous.

Table 3.17 Results of ANOVA and Tukey tests for the OEL bioassay data from the Tees and Mersey estuaries and the ranking of responses and ranking of responses

Site	Time	Variance ratio and level of probability				Ranking of responses at stations (lowest to highest)
		Time	Sample	Station	Time vs Station	
Tees Estuary	1-4	133.4 p<0.001	0.86 NS	1041 p<0.001	321.4 p<0.001	8,2,1,3,4,7,6,5
	1	-	-	1613 p<0.001	-	8,1,2,7,6,3,4,5
	2	-	-	1705 p<0.001	-	4,1=7,2,8,3,6,5
	3	-	-	1284 p<0.001	-	2,8,3,1,4,7,5,6
	4	-	-	1129 p<0.001	-	1,4,3,2,8,6,5=7
Mersey Estuary	1-3	28.8 p<0.001	1.08 NS	263.2 p<0.001	4.86 p<0.001	1,2,3,5,4,6,7,8
	1	-	-	232 p<0.001	-	1,5,2,7,6,3,4,8
	2	-	-	186.2 p<0.001	-	3,1,2,4,7,5,6,8
	3	-	-	355 p<0.001	-	1,4,2=5,3,6,7,8

Tisbe battagliai reproduction bioassay

Figures 3.20-3.23 show the results of *T. battagliai* reproduction bioassays (with juvenile production and lethality as endpoints) performed on four occasions for the Tees Estuary and three occasions for the Mersey Estuary. Water samples were collected at all eight stations at each site on each occasion.

Table 3.18 shows the results of the three and two way ANOVA and the ranking of responses between stations. ANOVA demonstrated that there were statistically significant differences between stations at both the Tees and Mersey estuaries using the *T. battagliai* lethality data when the sampling occasions were considered together and separately (see Table 3.18). Discrimination using the juvenile production data was only evident at the Tees and Mersey estuaries on some of the sampling occasions.

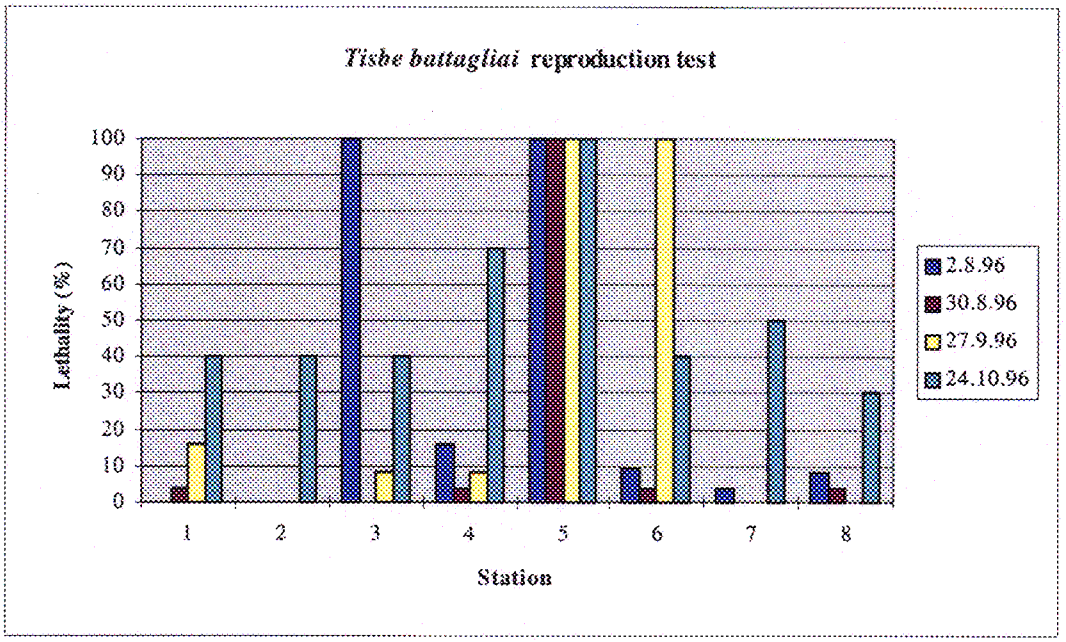

Figure 3.20 Mean *Tisbe battagliai* lethality on four sampling occasions for the Tees Estuary (* indicates no lethality measured)

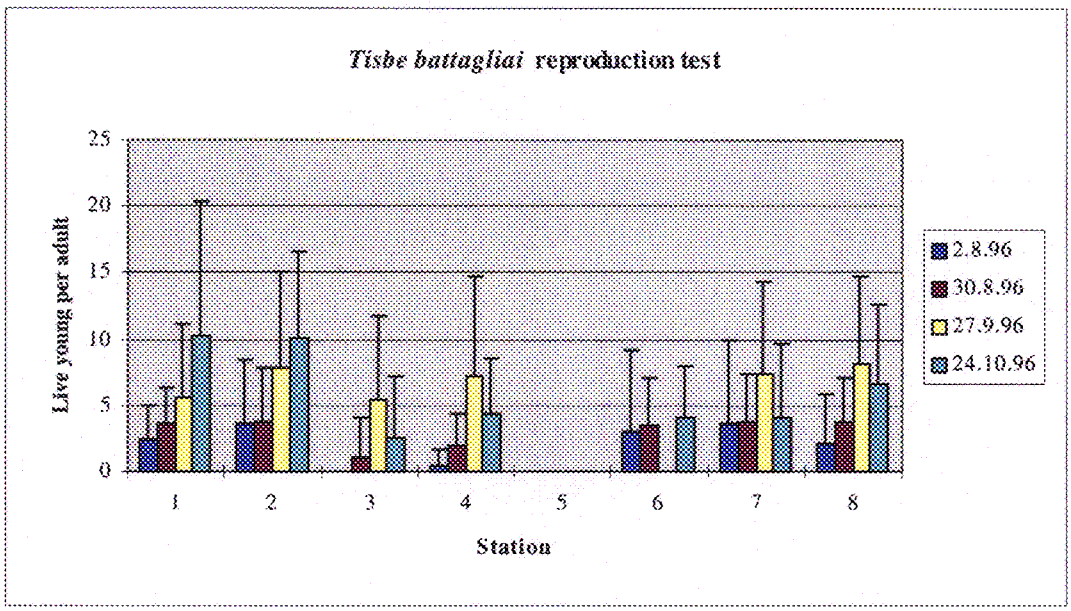

Figure 3.21 Mean *Tisbe battagliai* juvenile production on four sampling occasions for the Tees Estuary (error bars represent one standard deviation of station samples from the mean)

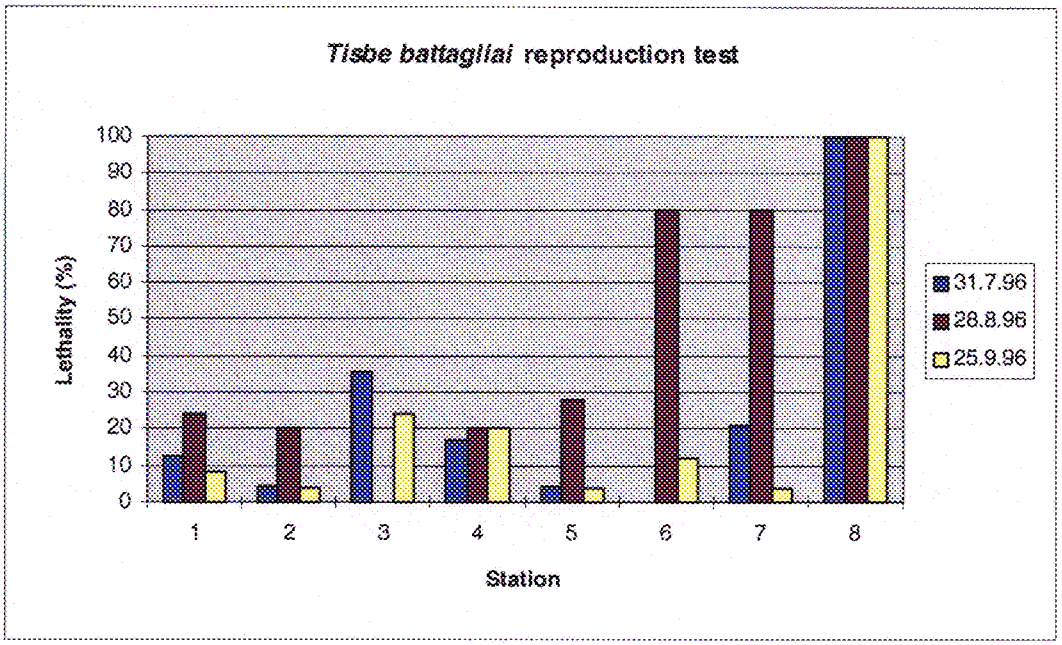

Figure 3.22 Mean *Tisbe battagliai* lethality on three sampling occasions for the Mersey Estuary (* indicates no lethality measured)

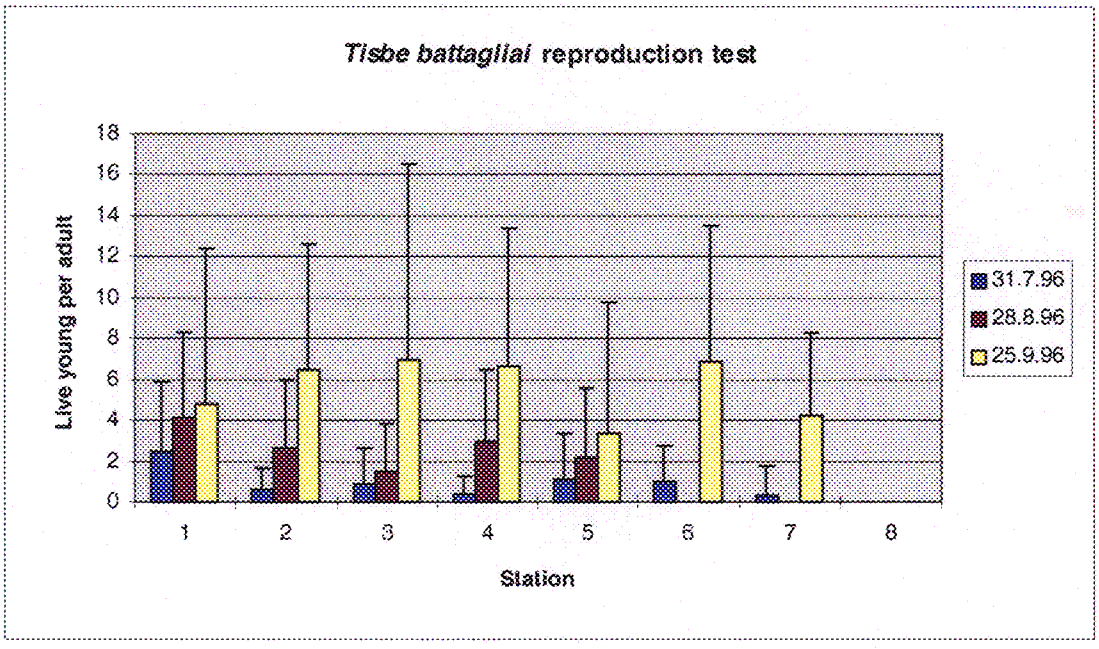

Figure 3.23 Mean *Tisbe battagliai* juvenile production on three sampling occasions for the Mersey Estuary (error bars represent one standard deviation of station samples from the mean)

Table 3.18 Results of ANOVA for the *Tisbe battagliai* reproduction bioassay data from the Tees and Mersey estuaries and ranking of responses

Site	Time	Variance ratio and level of probability				Ranking in responses at stations (lowest to highest for lethality, reverse for juvenile production)
		Time	Sample	Station	Time vs Station	
Tees Estuary Lethality	1-4	46.6 p<0.001	0.88 NS	83.7 p<0.001	14.8 p<0.001	2,8,7,1,4,3,6,5
	1	-	-	70.2 p<0.001	-	1,2,7,8,6,4,3=5
	2	-	-	124.8 p<0.001	-	2=3=7,1=4=6=8,5
	3	-	-	81.3 p<0.001	-	2=7=8,3=4,1,5=6
	4	-	-	3.35 p<0.001	-	8,1=2=3=6,7,4,5
Juvenile production	1-4	15.5 p<0.001	0.67 NS	5.82 p<0.001	2.5 p<0.01	2,8,1,7,4,6,3,5
	1	-	-	2.14 NS	-	2=7,6,1,8,4,3=5
	2	-	-	3.35 p<0.01	-	2=7=8,1,6,4,3,5
	3	-	-	0.65 NS	-	8,2,7,4,1,3,5=6
	4	-	-	9.3 p<0.001	-	1,2,8,4,6=7,3,5
Mersey Estuary Lethality	1-3	22.5 p<0.001	0.58 NS	52.4 p<0.001	8.66 p<0.001	2,5,1,4,3,6,7,8
	1	-	-	20.3 p<0.001	-	6,5,2,1,4,7,3,8
	2	-	-	31.2 p<0.001	-	3,2=4,1,5,6=7,8
	3	-	-	36.5 p<0.001	-	2=5=7,1,6,4,3,8
Juvenile production	1-3	74.3 p<0.001	3.27 p<0.05	2.0 NS	1.78 NS	1,2,4,6,3,5,7,8

Mean *T. battagliai* responses (juvenile production and lethality) measured in both the Tees and Mersey estuary samples were different for the different sampling occasions. The 3-way ANOVA showed these differences in sampling date to be statistically significant at both estuaries for juvenile production (Tees Estuary: variance ratio = 15.5, $p<0.001$, df=3, Mersey Estuary: variance ratio = 74.3, $p<0.001$, df=2) and lethality (Tees Estuary: variance ratio = 46.6, $p<0.001$, df=3, Mersey Estuary: variance ratio = 22.5, $p<0.001$, df=2). This may be due to tidal conditions at the sampling time and patterns in the discharge of contaminants and may include an element of spatial variability in the temporal variability.

On the Tees Estuary, the data showed complete *T. battagliai* lethality at Station 5 (the positive control) on all sampling occasions and at Stations 3, 4 and 6 on one occasion. At Stations 1, 2, 4, 7 and 8 lethality was low (≤16%) on the first three sampling occasion but considerably higher (>30%) at all these stations on the fourth sampling occasion. The number of live juveniles produced by surviving adults was always highest at Stations 1, 2, 7 and 8 compared to other stations.

On the Mersey Estuary lethality of *T. battagliai* was between 0 and 36% at Stations 1-5 on all occasions. At Stations 6 and 7 lethality was higher on the second sampling occasion (time 2), while at Station 8 there was complete lethality of organisms on all sampling occasions. The number of live juveniles produced by surviving adults showed no consistent pattern over time.

On the Tees Estuary, the complete lethality recorded at Station 5 on all occasions, Station 3 on the first sampling occasion and Station 6 on the third sampling occasion meant that no adults survived to produce juveniles. Therefore, these stations were excluded from the ANOVA of the juvenile production data. As a result juvenile production was compared at Stations 1, 2, 4, 6, 7 and 8 on the first sampling occasion, Stations 1, 2, 3, 4, 6, 7 and 8 on the second and fourth sampling occasions and Stations 1, 2, 3, 4, 7 and 8 on the third sampling occasion.

For the Mersey Estuary, the complete lethality recorded at Station 8 on all occasions and the high levels of lethality at Stations 6 and 7 on the second sampling occasion meant that no or few adults survived to produce juveniles. Therefore, these stations were excluded from the ANOVA of the juvenile production data. As a result juvenile production was compared at Stations 1, 2, 3, 4, 5, 6 and 7 on the first and third sampling occasions and Stations 1, 2, 3, 4 and 5 on the second sampling occasion.

In the bioassay lethality largely occurred after two days exposure in receiving water samples from both sites with only small increases in lethality (5-10%) occurring over the remaining seven days of the exposure period. Therefore a 48 h *T.battagliai* lethality test may be useful in receiving water monitoring..

The ANOVA also demonstrated that for both the Tees and Mersey estuaries, the differences between samples taken from the same stations were not generally significant (see Table 3.18) indicating a lack of spatial variability in the five spot water samples.

At both the Tees and Mersey Estuary sites the mean number of juveniles per surviving adult produced in the receiving water samples from all stations were lower than the mean number of 25.5 produced in the reference controls over the testing period (see Table 3.19). This finding indicates that at all stations where adult lethality was low the receiving water samples were having a sub-lethal effect on juvenile production.

Table 3.19 Summary of juvenile production in the *T. battagliai* reproduction tests at the different Tees and Mersey Estuary stations relative to that in seawater controls

Site	Time	Mean juvenile production in seawater	Mean juvenile production at each station as % of that in seawater control							
			1	2	3	4	5	6	7	8
Tees Estuary	1	26	9	14	-	15	-	12	14	8
	2	25	14	15	4	8	-	14	15	15
	3	31	18	25	17	24	-	-	24	26
	4	20	52	51	13	22	-	20	20	33
Mersey Estuary	1	26	9	2	3	1	4	4	1	-
	2	25	16	10	6	12	9	-	-	-
	3	31	16	21	22	22	11	22	14	-

3.1.4 Estuarine toxicity bioassays - sediment

Corophium volutator lethality bioassays

Figures 3.24 and 3.25 show the results of *C. volutator* lethality bioassays performed on two occasions for the Tees and Mersey estuaries respectively.

No sediment sample was taken at Station 3 on the second sampling occasion because tidal action had scoured the sediment.

Table 3.20 shows the results of the three and two way ANOVA and the ranking of responses between stations. ANOVA demonstrated that there were statistically significant differences between stations at both the Tees and the Mersey. This was the case when both sampling occasions were considered together and separately (see Table 3.20). However, as demonstrated by the station versus time interaction, the differences between stations for both estuaries were different for the two sampling occasions.

The 3-way ANOVA showed the differences in mean lethality with sampling date to be statistically significant for the two estuaries (Tees Estuary: variance ratio = 4.43, $p<0.05$, df=1, Mersey Estuary: variance ratio = 14.77, $p<0.001$, df=1 respectively). As with the freshwater sediment bioassays, these differences could be due to changes in weather conditions, discharge of contaminants and spatial variability. However, an additional factor potentially influencing estuarine sediment toxicity is tidal cycle, the bioassays results changing depending on whether the tide was rising or falling at the time of sampling. Despite the fact that all sampling was conducted as close as possible to low tide, the large number of stations studied meant that some were sampled before and some after low tide. One to two hours either way was usually regarded as acceptable given the logistical constraints of the study. One way of potentially reducing variability due to tidal cycle may be to conduct all sampling at a time when the tide is either rising and/or falling at all stations.

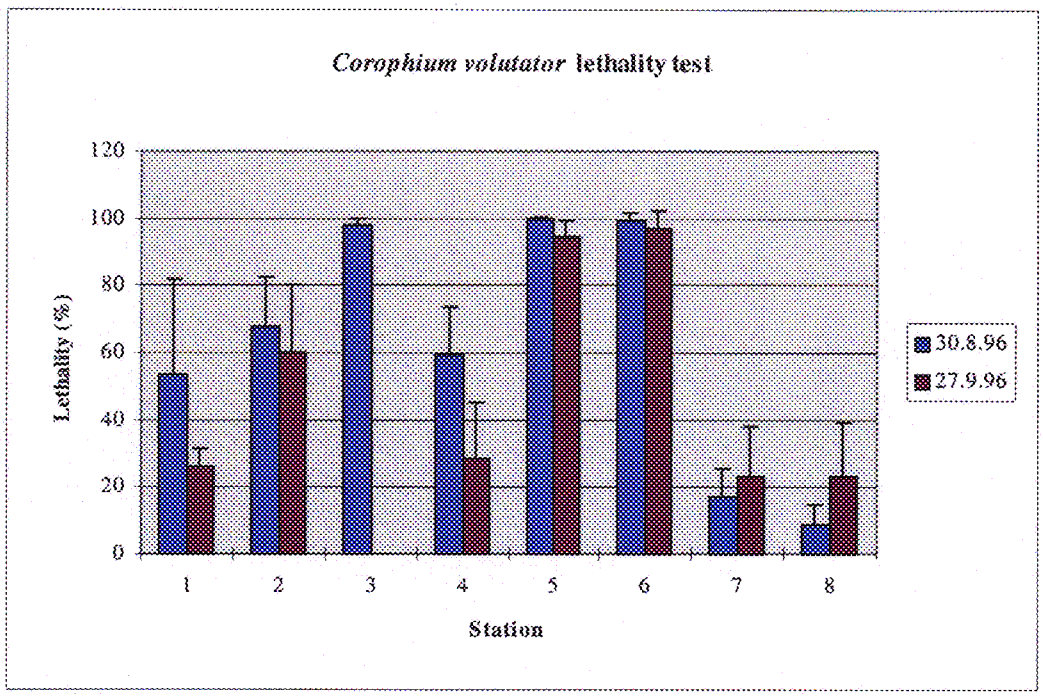

Figure 3.24 Mean lethality (10 d) of *Corophium volutator* on two sampling occasions for the Tees Estuary (error bars represent one standard deviation of station samples from the mean, NS = no samples taken)

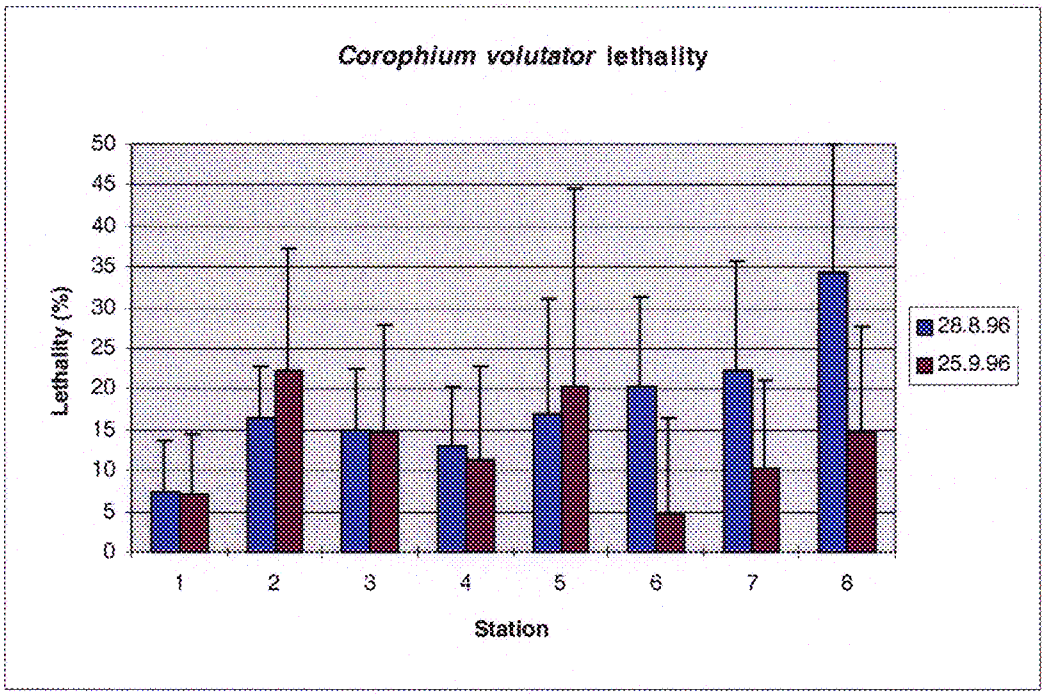

Figure 3.25 Mean lethality (10 d) of *Corophium volutator* on two sampling occasions for the Mersey Estuary (error bars represent one standard deviation of station samples from the mean)

Table 3.20 Results of ANOVA for the *Corophium volutator* lethality bioassay data from the Tees and Mersey estuaries and the ranking of responses

Site	Time	Variance ratio and level of probability				Ranking of responses at stations (lowest to highest)
		Time	Sample	Station	Time vs Station	
Tees Estuary	2 & 3	4.43 p<0.05	0.35 NS	71.1 p<0.001	3.61 p<0.05	8,7,1,4,2,5,6
	2	-	0.93 NS	84.2 p<0.001	-	8,7,1,4,2,3,6,5
	3	-	0.8 NS	34.3 p<0.001	-	8,7,1,4,2,5,6
Mersey Estuary	2 & 3	14.77 p<0.001	0.48 NS	5.13 p<0.001	3.98 p<0.01	1,4,6,3,7,5,2,8
	2	-	1.56 NS	6.62 p<0.001	-	1,4,3,2,5,6,7,8
	3	-	0.04 NS	3.68 p<0.01	-	6,2,7,4,3=8,5,2

For the Tees Estuary, the patterns of toxicity on both occasions were relatively consistent with the biological data for these stations. On both occasions sediments from Stations 3, 5, and 6 demonstrated the highest toxicity (>80% lethality) consistent with the fact that all three sites were situated in the immediate vicinity of the Dabholm Gut outfall (the positive control Station 5 in the mixing zone, Station 6 immediately downstream and Station 3 immediately upstream). Both sampling times identified Station 2 sediments, which were situated on the bank immediately opposite the outfall, as having moderate toxicity (≥ 60% lethality). In addition, on both sampling occasions Stations 7 and 8 were identified as having the least toxic sediments (<25% lethality). This is consistent with the fact that these were considered as unpolluted sites with Station 8 being regarded as the reference location based on historical data (Environment Agency, unpublished observations).

Table 3.21 summarises the results of 10d *C. volutator* lethality tests carried out on sediment samples collected at the Tees Estuary stations over the period 1994-1996. Sediments from Stations 7 and 8 (the reference location) always showed low levels of lethality (<27% in all cases) while those from Stations 5 (the positive control) and 6 always showed high levels of lethality (≥75%) on all occasions. The levels of lethality in sediments from other stations varied over time.

Table 3.21 Summary of data from 10d *C. volutator* lethality tests carried out on sediments from the Tees Estuary

Station	Mean % lethality in 10 d *C. volutator* tests carried on different occasions			
	EA - 1994[1]	EA 1995[1]	WRc 1996 (1 - this study)	WRc 1996 (2 - this study)
1	ND	0	53.5	26.3
2	ND	100	67.6	20.2
3	35	100	98.2	ND
4	ND	100	59.5	28.7
5	75	100	99.6	94.3
6	100	100	99	97
7	13.3	1.6	17.3	23.3
8	6.6	26.7	8.9	23

1 Unpublished observations

For the Mersey, no marked level of lethality (>20%) was seen at any of the stations with the exception of Stations 7 and 8 on the first sampling occasion. Even at these stations the percentage lethality was below 35%. The lack of differences in effects between the stations is likely to reflect the fact that the sediments were not lethally toxic to *Corophium* over a 10 day period. For this reason, a longer term (28 day) *Corophium* growth and lethality bioassay was performed on samples taken from this estuary on the third sampling occasion.

The ANOVA demonstrated that for both the Tees and the Mersey, differences between spot samples taken from within the same stations were not significant in influencing the patterns of toxicity observed between stations (see Table 3.20). This is consistent with the results of the freshwater 10 day *C.riparius* lethality bioassay where spatial variability did not seem to influence bioassay responses within stations. Tables 3.22 and 3.23 show summary information for the physico-chemical determinands measured in spot samples for the two sampling occasions.

By observing the standard deviation of each physico-chemical determinand at each site it is clear that there was generally a limited amount of variability (<25%) in sediment type between spot samples. However, this variation was not sufficient to cause significant differences in the toxic response between the samples.

However, as discussed for the freshwater sediment bioassays, although spatial variability was not a major factor influencing bioassay response within stations, it may have been contributing to the overall pattern of toxicity observed between stations. Table 3.24 shows the ranges in physico-chemical determinands which have been demonstrated to support survival of *Corophium* in previous studies.

Table 3.22 Summary of physico-chemical determinands measured in Tees Estuary sediments on both sampling occasions

Time -station	TOC (%)		Particle size (<63 μm) (%)		Water content (%)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
1-1	8.0	1.2	84.4	4.6	ND	ND
1-2	3.0	1.2	26.0	6.5	ND	ND
1-3	3.9	0.1	45.6	3.6	ND	ND
1-4	3.0	1.1	48.0	28.2	ND	ND
1-5	18.5	7.6	23.2	7.5	ND	ND
1-6	8.3	0.6	51.4	2.7	ND	ND
1-7	1.6	0.7	6.8	2.0	ND	ND
1-8	3.6	0.8	35.1	6.9	ND	ND
2-1	8.8	0.7	77.4	5.3	56.9	4.5
2-2	4.3	0.7	42.6	4.2	39.3	3.7
2-3	-	-	-	-	-	-
2-4	0.9	0.2	9.3	2.7	23.4	0.5
2-5	12.7	1.4	48.6	6.9	41.8	2.3
2-6	-	-	-	-	-	-
2-7	1.7	0.4	7.4	2.5	25.2	0.7
2-8	2.7	0.6	41.7	16.9	30.0	3.0

Table 3.23 Summary of physico-chemical determinands measured in Mersey Estuary sediments on both sampling occasions

Time -station	TOC (%)		Particle size (<63 μm) (%)		Water content (%)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
1-1	0.3	0.1	87.1	2.6	ND	ND
1-2	1.3	0.4	41.7	7.7.	ND	ND
1-3	0.6	0.1	68.5	5.7	ND	ND
1-4	1.6	0.6	43.5	14.0	ND	ND
1-5	1.2	0.3	48.2	8.5	ND	ND
1-6	2.1	0.2	17.4	2.5	ND	ND
1-7	1.3	0.1	47.4	4.4	ND	ND
1-8	1.2	0.4	42.4	10.3	ND	ND
2-1	0.6	0.1	70.9	7.4	32.0	2.7
2-2	6.6	0.1	69.6	3.2	31.0	1.9
2-3	1.3	0.3	37.8	6.7	43.2	6.3
2-4	1.2	0.3	53.2	4.5	35.0	2.4
2-5	1.0	0.2	52.7	5.2	35.6	3.7
2-6	2.0	0.1	19.2	2.6	49.3	1.9
2-7	1.4	0.3	47.9	7.5	34.4	6.0
2-8	1.6	0.1	35.0	3.3	44.4	2.3

Table 3.24 Ranges of physico-chemical factors which support 10 day *Corophium* survival

Sediment factor	Range (%)	Sediment type	Test duration	Reference
TOC	0.47-0.87	natural	10 d	Fleming and Sims (1997)
	4.4	natural	10 d	" " "
			28 d	
	4.1-16.3	natural	10 d	Fleming *et al* (1993)
			28 d	
< 63μm fraction	7-97	natural	28 d	Fleming and Sims (1997)
Water content	30-33	natural	10 d	Fleming and Sims (1997)

A comparison of these ranges with the mean values for test sediments in Tables 3.22 and 3.23 shows that most of the test sediment characteristics fall within ranges previously demonstrated to support *Corophium* survival. Again, insufficient data are available to perform this comparison for water content. The one notable exception is the TOC content in Station 5 sediment on the Tees. The highest measurement of 18.5% in sediment from this station lies outside the range previously tested. However, this high value is likely to be a direct result of the industrial discharge at this station as the TOC of all other sediments in the estuary fall below 9%. Therefore, this increased TOC should be regarded as a stressor in itself and as any effects due to this parameter would be considered as adverse it should not be separated from toxicant impact.

Table 3.25 shows the results of analyses of correlations between the 10 day lethality data and sediment physico-chemical parameters. Few significant correlations were observed between TOC, particle size or water content and 10 d lethality. The exceptions were on the Tees Estuary on the second sampling occasion where a significant correlation was seen between lethality and TOC ($r=0.67$, $p<0.001$, $n=26$) and on the first sampling occasion on the Mersey Estuary where a significant correlation was seen between lethality and particle size (<63μm, $r=0.35$, $p<0.05$, $n=46$). For the Tees, the correlation between 10 day *C.volutator* lethality and TOC was positive showing that lethality increased as TOC increased. This is contrary to the results of freshwater 10 d *C.riparius* lethality bioassays at the Willow Brook where a significant negative correlation between TOC and lethality was observed on one occasion. The correlation for the Tees could be due to an intercorrelation with TOC and levels of toxicants in the sediment as a result of the Station 5 discharge. The TOC was highest at this station with a corresponding high level of lethality. For the Mersey Estuary, the correlation with particle size was also positive, demonstrating that as silt/clay fraction increased, lethality also increased. Whether these correlations are linked with toxicant concentrations or due to natural stress of sediment type is not possible to differentiate further without analysis of toxicant concentrations in the sediments and a thorough investigation of physico-chemical influences. However, the comparison with previous sediment type data would suggest that the sediment types encountered in this study should have supported the survival of *Corophium* if they were uncontaminated, suggesting that the pattern of response between stations can be largely attributed to either toxicant concentrations or to other unidentified sediment factors which were not measured in this study.

Table 3.25 Correlation coefficients (and significance) between 10 day lethality data and sediment physico-chemical parameters

Site	Time	Correlation coefficients (and significance) between lethality data and physico-chemical parameters		
		TOC	Particle size (<63 μm)	Water content
Tees Estuary	2	0.104 (NS)	0.027 (NS)	ND
	3	0.670 (p<0.001)	0.220 (NS)	0.241 (NS)
Mersey Estuary	2	0.283 (NS)	0.351 (p<0.05)	ND
	3	-0.366 (NS)	-0.246 (NS)	-0.290 (NS)

ND = no data, NS = not significant

Corophium volutator growth/lethality bioassay

Figures 3.26-3.29 show the results of the *Corophium volutator* growth/lethality bioassay performed on two occasions. The first two figures show lethality and growth respectively for the Tees Estuary on the first test occasion (time 2) . The second test for the Tees Estuary sediments failed and the results have not been given. Although this test was a non-core test and intended for the Tees Estuary only, as mentioned above it was decided to also perform a test for the Mersey Estuary sediments, as this was the site at which longer term sub-lethal sediment toxicity was more likely to be observed. Lethality was seen in the 10 day *C.volutator* bioassay at most stations on both sampling occasions on the Tees Estuary.

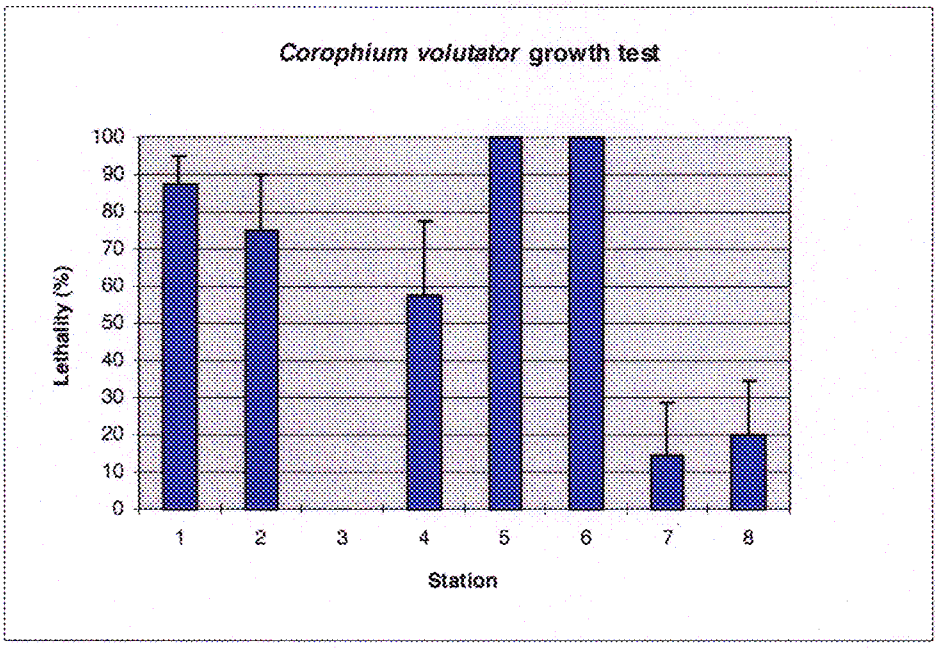

Figure 3.26 Mean lethality (28 d) of *Corophium volutator* for the Tees Estuary (error bars represent one standard deviation of station samples from the mean, NS = no samples taken)

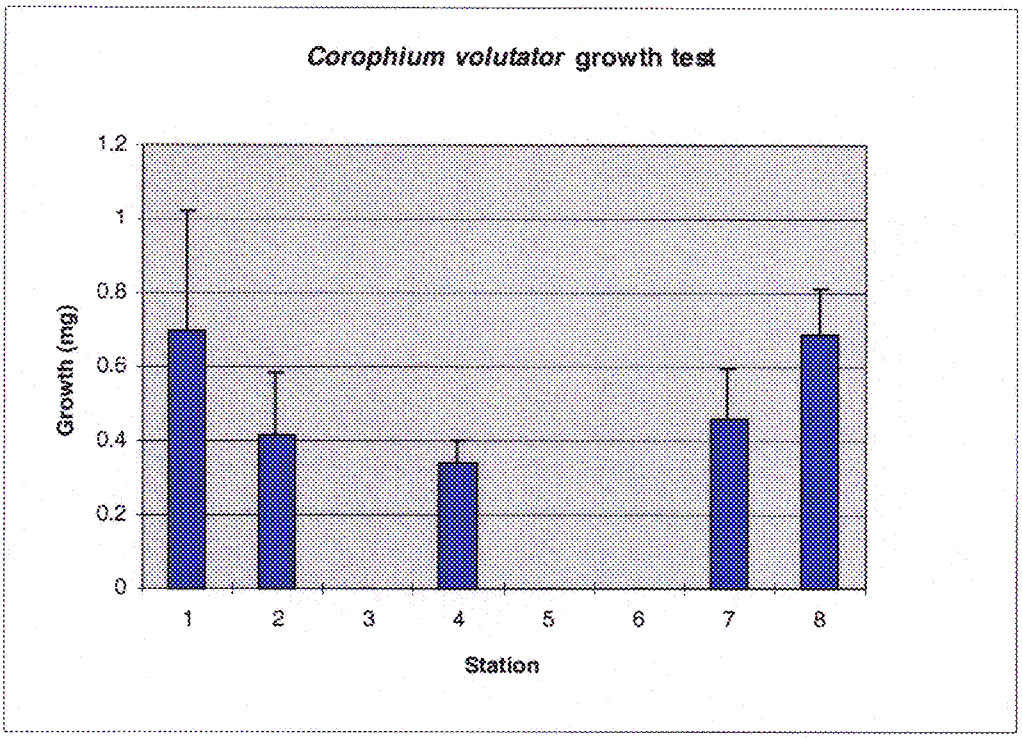

Figure 3.27 Mean growth (as dry weight) of *Corophium volutator* for the Tees Estuary (error bars represent one standard deviation of station samples from the mean, NS = no samples taken)

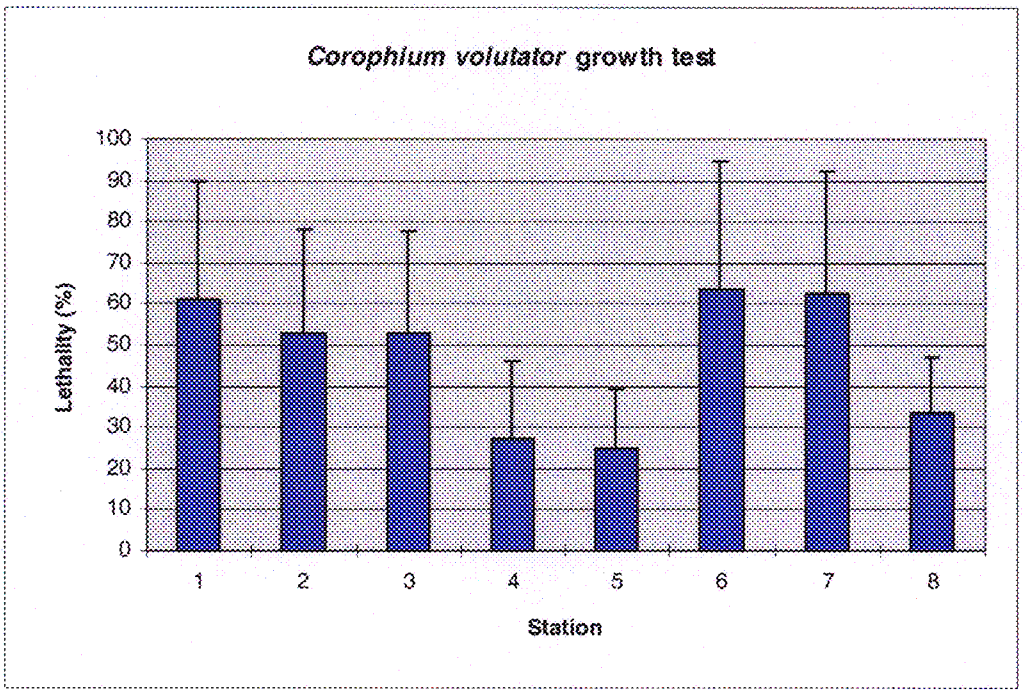

Figure 3.28 Mean lethality (28 d) of *Corophium volutator* for the Mersey Estuary (error bars represent one standard deviation of station samples from the mean)

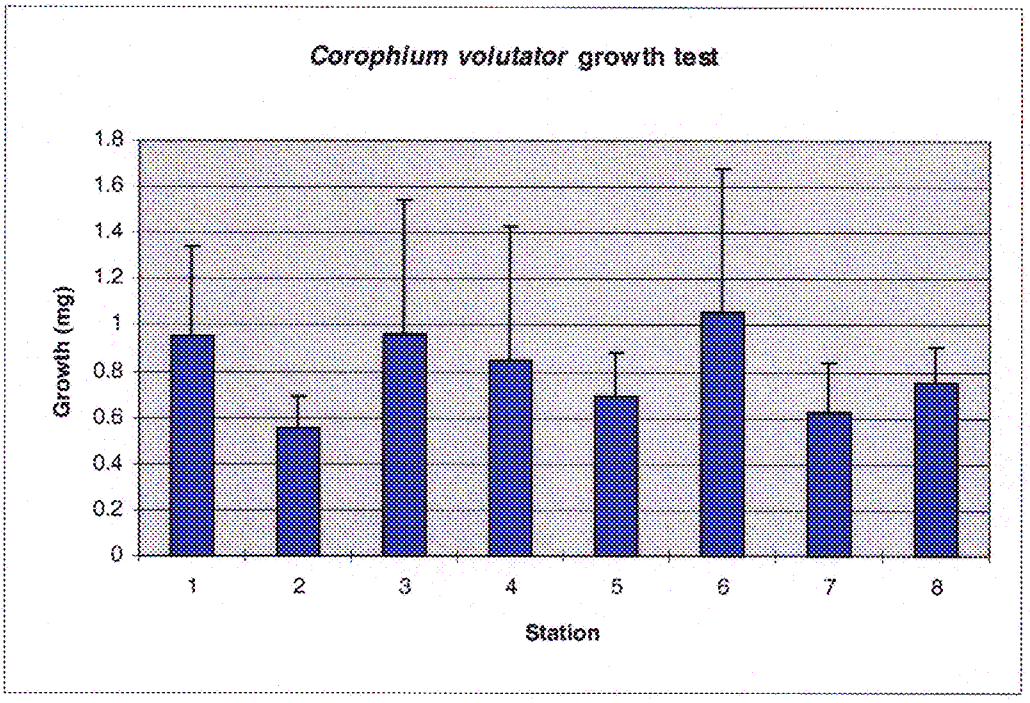

Figure 3.29 Mean growth (as dry weight) of *Corophium volutator* for the Mersey Estuary (error bars represent one standard deviation of station samples from the mean)

In sediment from Tees Estuary Stations 5 (the positive control) and 6, 100% lethality was observed and so no survivors were available for measurement of the growth endpoint.

Tables 3.26 and 3.27 show the results of the three and two way ANOVA and the ranking of responses for the Tees and Mersey estuaries respectively.

Table 3.26 Results of ANOVA for the *Corophium volutator* growth/ lethality bioassay data from the Tees Estuary and the ranking of responses

Effect	Variance ratio and level of probability					Ranking of responses at stations (lowest to highest for lethality, reverse for growth)
	Time	Time	Sample	Station	Time vs Station	
Lethality	2	-	4.01 p<0.01	196.4 p<0.001	-	7,8,4,2,1,5=6
	3	Test failed				
Growth	2	-	5.75 p<0.001	29.7 p<0.001	-	1,8,7,2,4
	3	Test failed				

Table 3.27 Results of ANOVA for the *Corophium volutator* growth/ lethality bioassay data from the Mersey Estuary and the ranking of responses

Effect	Time	Variance ratio and level of probability Time	Sample	Station	Time vs Station	Ranking of responses at stations (lowest to highest for lethality, reverse for growth)
Lethality	2		Not tested			
	3	-	2.29 NS	7.21 p<0.001	-	5,4,8,2=3,1,7,6
Growth	2		Not tested			
	3	-	0.88 NS	3.38 p<0.01	-	6,3,1,4,8,5,7,2

For the Tees Estuary, lethality over 28 days showed a similar pattern to that observed in the 10 day lethality bioassays performed on the first sampling occasion. The sensitivity of the 28 day lethality endpoint was either equivalent to or greater than 10 day lethality. ANOVA demonstrated that differences between stations were significant in determining the pattern in the 28 day lethal response (variance ratio =196, $p<0.001$, df=6).

However, ANOVA also demonstrated that differences between samples taken from the same station, that is spatial variability, were also influential in determining the pattern of toxic response (variance ratio = 4.01, $p<0.01$, df=4)(see Table 3.26). Because differences were observed within sites, the influence of sediment type in determining patterns between stations is likely to be more pronounced than in the 10 day lethality bioassays. Therefore, the influence of spatial variability on these endpoints should be carefully considered if the tests are used for receiving water monitoring. A regression model to normalise the effects of sediment type may be one way of overcoming this problem.

Lethality after 28 days was significantly correlated with TOC ($r=0.6$, $p<0.001$, $n=28$) and particle size (<63μm, $r=0.53$, $p<0.05$, $n=28$). Both factors were positively correlated which demonstrates that lethality increased as silt/clay fraction and TOC increased. The influence of sediment type versus toxicant concentration cannot be identified further without analysis of toxicant concentrations and an investigation of the influence of sediment type.

For the Mersey Estuary, lethality over 28 days gave a significant response (that is >20% lethality) at all stations thus demonstrating greater sensitivity than the 10 day bioassay performed on the same samples (see Figure 3.25). ANOVA demonstrated that there were statistically significant differences between stations in the 28 day bioassay (variance ratio = 7.21, $p<0.001$, df=7). The lack of further discrimination is likely to be due to the toxicity of the sediments, all of which showed <65% lethality. ANOVA showed that for the Mersey Estuary, spatial variability within stations was not influential in determining the overall pattern of toxic response (see Table 3.27). Lethality after 28 days was not significantly correlated with either TOC, particle size (<63 μm) or water content.

For the 28 day growth endpoint on the Tees Estuary samples, ANOVA demonstrated that there were statistically significant differences between stations (variance ratio = 30.4, $p<0.001$, df=4). The growth endpoint was not correlated with lethality which raises questions about the biological significance of this endpoint. For the Tees sediments, growth did not add any further discrimination to the lethality endpoint. For the Mersey sediments, in which less lethality was observed overall, the growth endpoint did discriminate between stations and was negatively correlated with lethality ($r=-0.43$, $p<0.05$) as would be expected if growth was a biologically significant measure of adverse effect. It may be that the growth endpoint is more useful when a lower level of lethality is observed. However, when a certain level of lethality occurs in the test system the growth endpoint becomes more difficult to interpret in terms of biological significance, and more influenced by the design of the test systems and sediment factors such as nutritive quality.

For both the Tees and Mersey Estuary samples the growth endpoint was significantly correlated with TOC (Tees $r=0.491$, $p<0.05$, $n=25$; Mersey $r=0.346$, $p<0.05$, $n=37$) but not particle size.

Arenicola marina cast formation bioassay

Figures 3.30 and 3.31 show the results of the *Arenicola* cast formation bioassay performed on two occasions on the Tees Estuary. Table 3.28 shows the results of the three and two way ANOVA and the ranking of responses between stations.

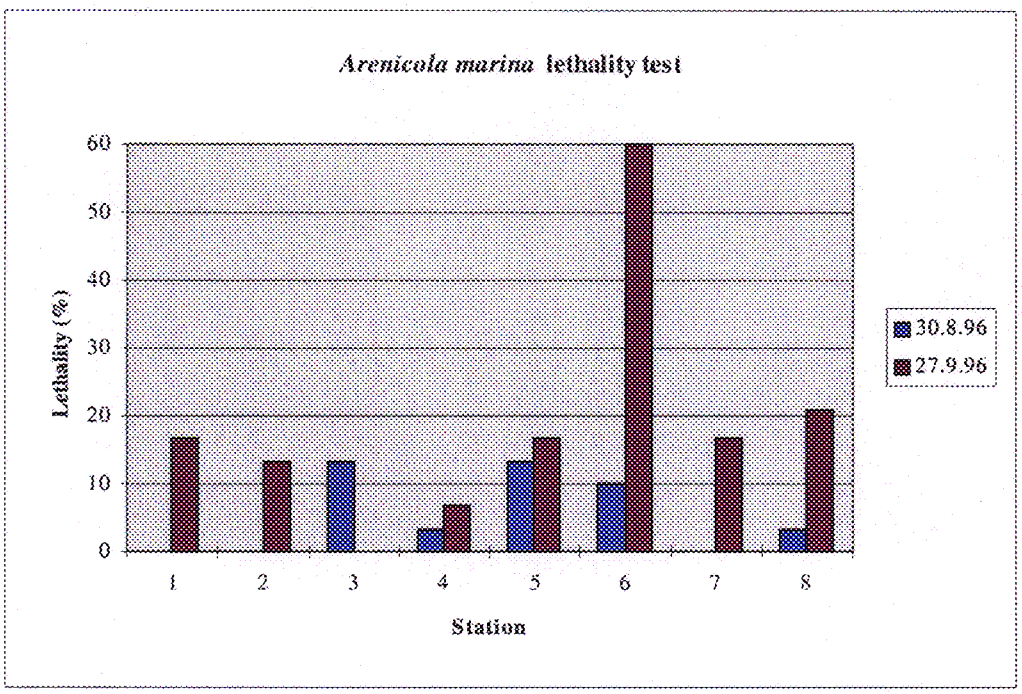

Figure 3.30 Mean lethality of *Arenicola marina* on two sampling occasions for the Tees estuary (error bars represent one standard deviation of station samples from the mean . NS = no samples taken, * indicates no lethality measured)

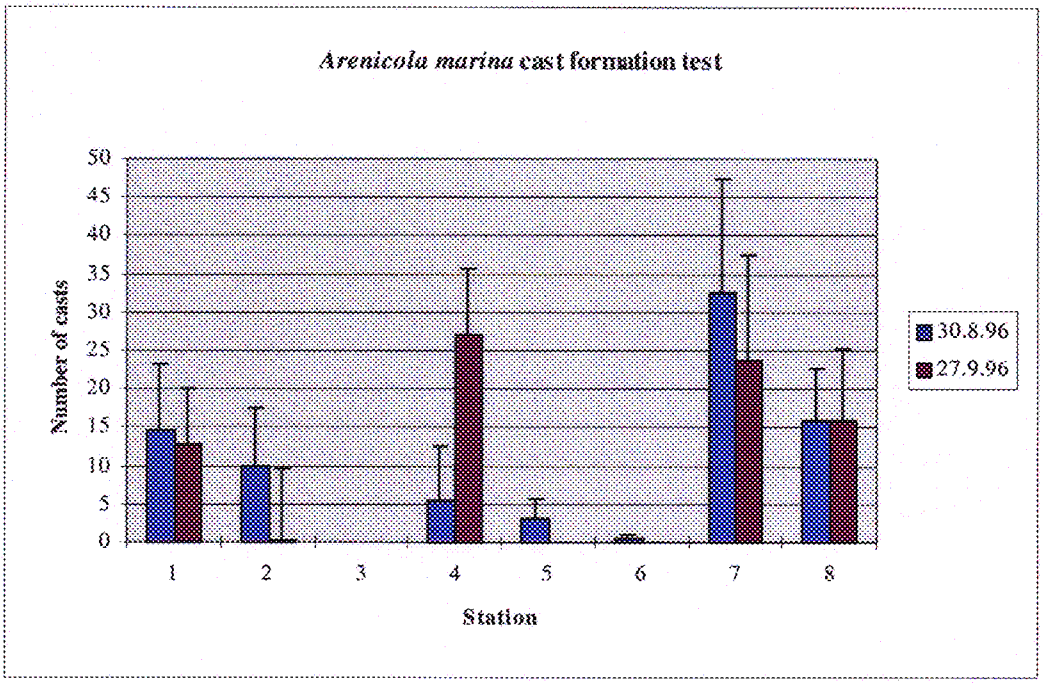

Figure 3.31 Mean cast formation of *Arenicola marina* on two sampling occasions for the Tees estuary (error bars represent one standard deviation of station samples from the mean)

Table 3.28 Results of ANOVA for the *Arenicola marina* cast formation/lethality data from the Tees Estuary and the ranking of responses

Effect		Variance ratio and level of probability				Ranking of responses at stations (lowest to highest for lethality, reverse for cast formation)
	Time	Time	Sample	Station	Time vs Station	
Lethality	2 & 3	38.2 p<0.001	3.43 p<0.01	3.84 p<0.01	2.06 NS	4,2,1=7,8,5,6
	2	-	-	0.98 NS	-	1=2=7,4=8,6,3=5
	3	-	- NS	2.93 p<0.05	-	4,2,1=5=7,8,6
Cast formation	2 & 3	0.1 NS	2.19 NS	23.3 p<0.001	6.37 p<0.001	7,8,4,1,2,5,6
	2	-	-	10.85 p<0.001	-	7,8,1,2,4,5,6,3
	3	-	-	12.32 p<0.01	-	4,7,8,1,2,5=6

Mean lethality of *Arenicola* was different for the two sampling occasions. The 3-way ANOVA showed this difference to be significant (variance ratio = 38.2, $p<0.001$, df=1). As with the *Corophium* bioassays, these differences could be due to changes in weather conditions, discharge of contaminants, spatial variability or tidal cycle.

ANOVA also demonstrated that differences between stations were also significant in determining the patterns in lethal toxic response when both sampling occasions were considered together, but when considered separately, the first sampling occasion showed no significant effect of station (see Table 3.28). Given that lethality was just over or below 20% at all other sampling stations on both sampling occasions, the increased lethality at Station 6 on the second sampling occasion is likely to have been an anomaly. This is supported by the fact that the sediment from this station did not cause significant lethality on the first occasion, and Station 5 sediment (the positive control), which has been shown to be as toxic as Station 6 in all other tests, did not lead to lethality in this test. The lack of sensitivity and discrimination in *Arenicola* lethality is not a concern as the primary objective of this bioassay is to determine the toxic response using the sub-lethal endpoint of cast formation.

For the cast formation endpoint, differences between sampling occasions can be seen from Figure 3.31 although the 3-way ANOVA showed that temporal variability was not significant in determining patterns when compared with the station effect (see Table 3.28). However, the station versus time interaction showed that the differences between stations were different for the two sampling occasions.

Unlike the 28 day *Corophium* growth/lethality bioassay, differences between samples taken from the same stations did not significantly influence the pattern of toxicity observed between stations for *Arenicola* cast formation (see Table 3.28). This is consistent with the 10 day *Corophium* lethality bioassay. However, differences in sediment type may still be influencing the patterns in cast formation observed between stations. Previous testing with *Arenicola* at WRc has shown that both lethality and cast formation are supported within the following particle size range: 7-97% as <63μm fraction. Insufficient data are available within house to compare the other physico-chemical characteristics of the test sediments although this information may to have been generated by MAFF. The test sediments from the Tees all fall within this particle size range. For *Arenicola* lethality there were no significant correlations with TOC, particle size, or water content (see Table 3.29). For cast formation significant correlations were evident on the second sampling occasion where negative correlations were seen with TOC ($r=-0.6$, $p<0.001$, $n=26$) particle size (<63 μm, $r=0.504$, $p<0.01$, $n=26$) and water content ($r= -0.503$, $p<0.01$, $n=26$). This is similar to the 10 day *Corophium* lethality bioassay for which a correlation with TOC was observed on the second sampling occasion. The fact that this correlation was observed for both species suggests that the toxic response of both was mediated by intercorrelation of TOC with toxicant concentrations rather than the specific test organism requirements. The influence of sediment type versus toxicant concentrations can only be differentiated further by analysis of toxicant concentrations in the sediments and a thorough investigation of sediment type.

Table 3.29 Correlation coefficients (and significance) between 10 day cast formation and lethality data in the *Arenicola marina* bioassay and sediment physico-chemical parameters

Endpoint	Time	Correlation coefficients (and significance) between lethality data and physico-chemical parameters		
		TOC	Particle size (<63 μm)	Water content
Lethality	2	0.358 (NS)	0.048 (NS)	ND
	3	0.206 (NS)	0.153 (NS)	0.26 (NS)
Cast formation	2	-0.361 (NS)	-0.352 (NS)	ND
	3	-0.601 (p<0.001)	-0.504 (p<0.01)	-0.503 (p<0.01)

ND = no data, NS = not significant

3.1.5 *In situ* bioassays - freshwater

***Gammarus pulex in situ* (feeding rate) bioassay**

Figures 3.32-3.35 show the results of *G. pulex in situ* bioassays (with feeding rate and lethality as endpoints) performed on two occasions for the River Aire and the Willow Brook. Cages could not be deployed at Stations 3 and 7 on the River Aire and at Station 3 on the Willow Brook. Table 3.30 shows the results of the three and two way ANOVA and the ranking of responses between stations. ANOVA on leaf correction factors and *Gammarus* weight indicated these factors did not vary between stations.

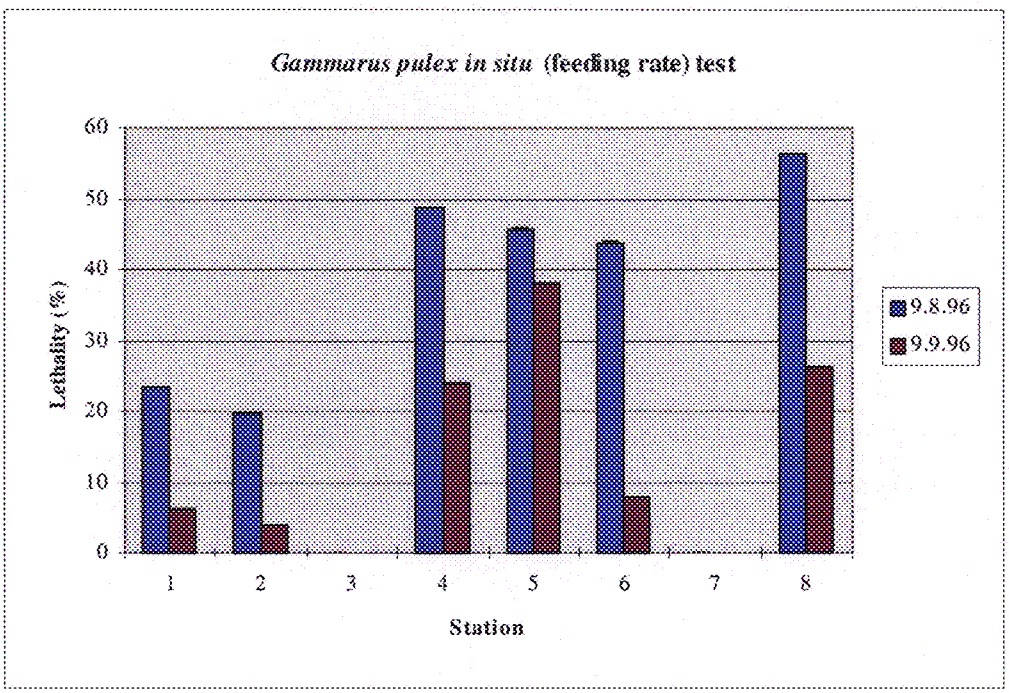

Figure 3.32 Mean lethality of *Gammarus pulex* during two *in situ* deployments in the River Aire (ND = not deployed)

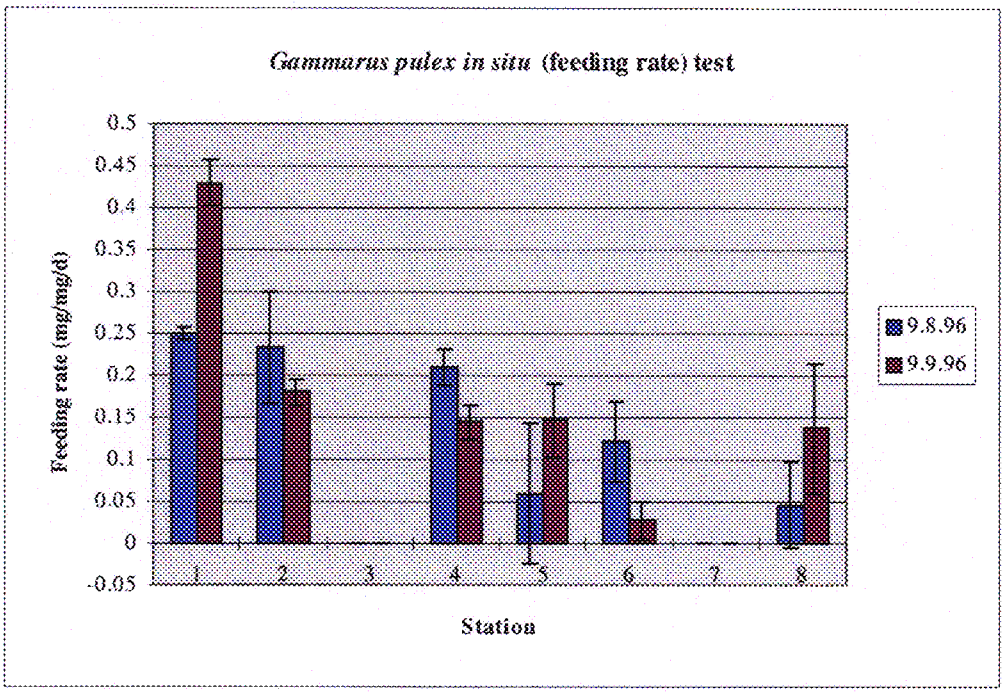

Figure 3.33 Mean feeding rate of *Gammarus pulex* during two *in situ* deployments in the River Aire (error bars represent one standard deviation of cage replicates from the mean, ND = not deployed)

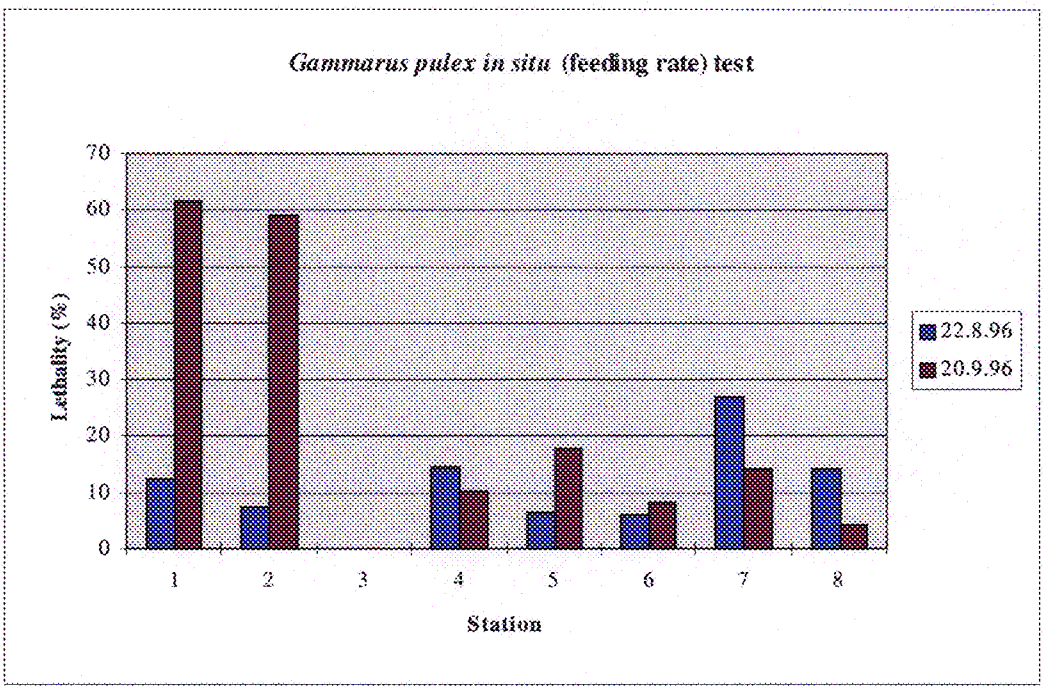

Figure 3.34 Mean lethality of *Gammarus pulex* during two *in situ* deployments in the Willow Brook (ND = not deployed)

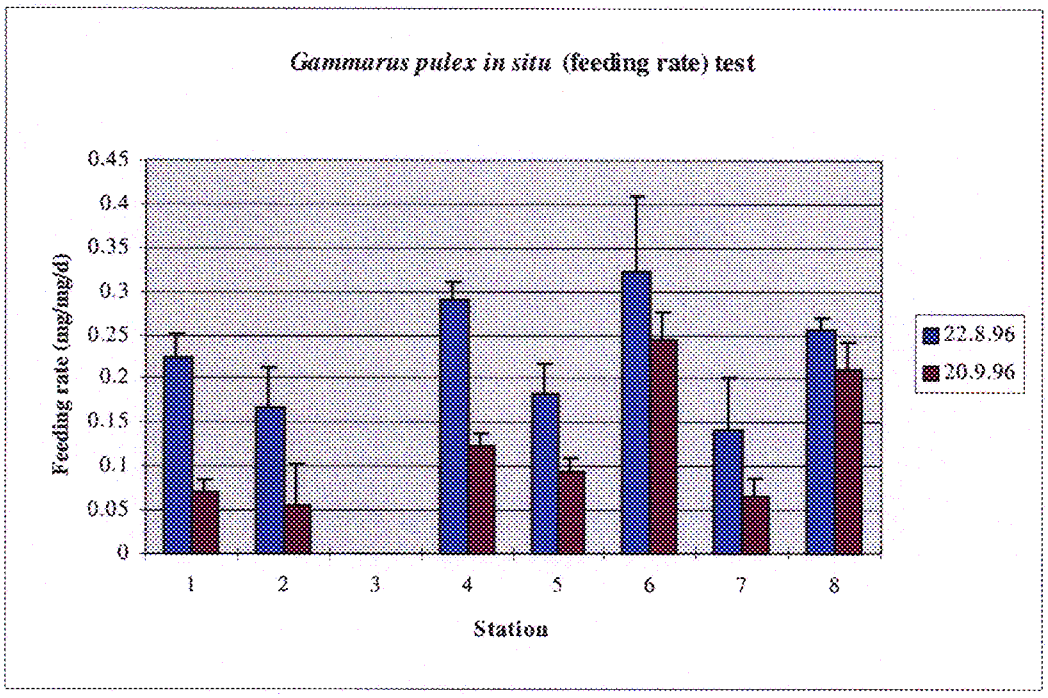

Figure 3.35 Mean feeding rate of *Gammarus pulex* during two *in situ* deployments in the Willow Brook (error bars represent one standard deviation of cage replicates from the mean, ND = not deployed)

ANOVA demonstrated that there were differences between stations at both the River Aire and the Willow Brook. This was the case when both deployment occasions were considered together and separately (see Table 3.30). However, as demonstrated by the station versus time interaction, the differences between stations for both rivers was different for the two deployment occasions.

The 3-way ANOVA showed the differences in mean *G. pulex* feeding rate and lethality with deployment date to be statistically significant for both lethality at the River Aire (variance ratio = 59.6, $p<0.001$, df=1) and feeding rate at the Willow Brook (variance ratio = 77.0, $p<0.001$, df=1) respectively. This may be due to flow conditions up to and including the sampling time and patterns in the discharge of contaminants.

On the River Aire, the first deployment (time 2) was immediately following a rain event whereas the second deployment (time 3) was at a period of low flow. The data showed that lethality was always lower at Stations 1 and 2 (<20%) relative to Stations 4, 5, 6 and 8, though lethality was higher at all stations on the first deployment than the second. The pattern of response for feeding rate also varied over time, though animals at Station 8 had lower feeding rates than Stations 1 ,2 and 4 on both occasions. Despite the high mortality of *G. pulex* at Station 4 on the first deployment (51 %), surviving animals showed comparable feeding rates to animals at the upstream stations (1 and 2).

Table 3.30 Results of ANOVA for the *G. pulex in situ* bioassay data from the River Aire and Willow Brook and ranking of responses

Site	Time	Variance ratio and level of probability				Ranking of responses at stations (lowest to highest for lethality, reverse for feeding rate)
		Time	Cage	Station	Time vs Station	
River Aire Lethality	2 & 3	59.6 p<0.001	0.83 NS	8.79 p<0.001	1.92 NS	1,2,6,,4,8,5
	2	-	-	4.31 p<0.05	-	1,2,6,5=8,4
	3	-	-	4.98 p<0.05	-	1=2,6,4,8,5
Feeding rate	2 & 3	3.93 NS	0.03 NS	11.5 p<0.001	6.15 p<0.05	1,2,4,6,5,8
	2	-	-	24.0 p<0.001	-	4,2,1,6,8,5
	3	-	-	67.9 p<0.001	-	1,2,5,4,8,6
Willow Brook Lethality	2 & 3	3.74 NS	0.4 NS	2.84 NS	1.18 NS	8=6,4,5,7,1,2
Feeding rate	2 & 3	77.0 p<0.001	0.44 NS	20.1 p<0.001	3.65 p<0.05	6,8,4,1,5,2,7
	2	-	-	6.11 p<0.001	-	6,4,8,1,5,2,7
	3	-	-	16.9 p<0.001	-	6,8,4,1,5,2,7

The feeding rates measured at Station 1 on the two deployments (0.27-0.43 mg mg^{-1} d^{-1}) were similar to the values of (0.17-0.19 mg mg^{-1} d^{-1}) at two sites upstream of this station in a previous study (Crane *et al* 1993), accepting that a different population of *Gammarus* was used and there may have been temperature differences between the stations.

On the Willow Brook no obvious differences in flow conditions were noted at the time of deployment and recovery. However, this does not preclude the influence of weather events and flow regime in the deployment. The data showed that lethality was generally low (<20%) at all stations with the exception of Stations 1 and 2 on the second deployment occasion (time 3). The pattern of response for feeding rate varied over time though values were higher at Stations 6 and 8.

Daphnia magna in situ (lethality) bioassay

Figure 3.36 shows the results of the *Daphnia in situ* bioassays performed on two occasions for the River Aire. On both occasions the tests were deployed for six days. Originally a two day

deployment had been planned but preliminary work showed that extending the exposure duration improved discrimination between stations without leading to a significant decrease in survival at less impacted stations. Cages were not deployed at Station 7 as access was not possible and cages were lost at Station 6 on the second sampling occasion (time 3) due to vandalism.

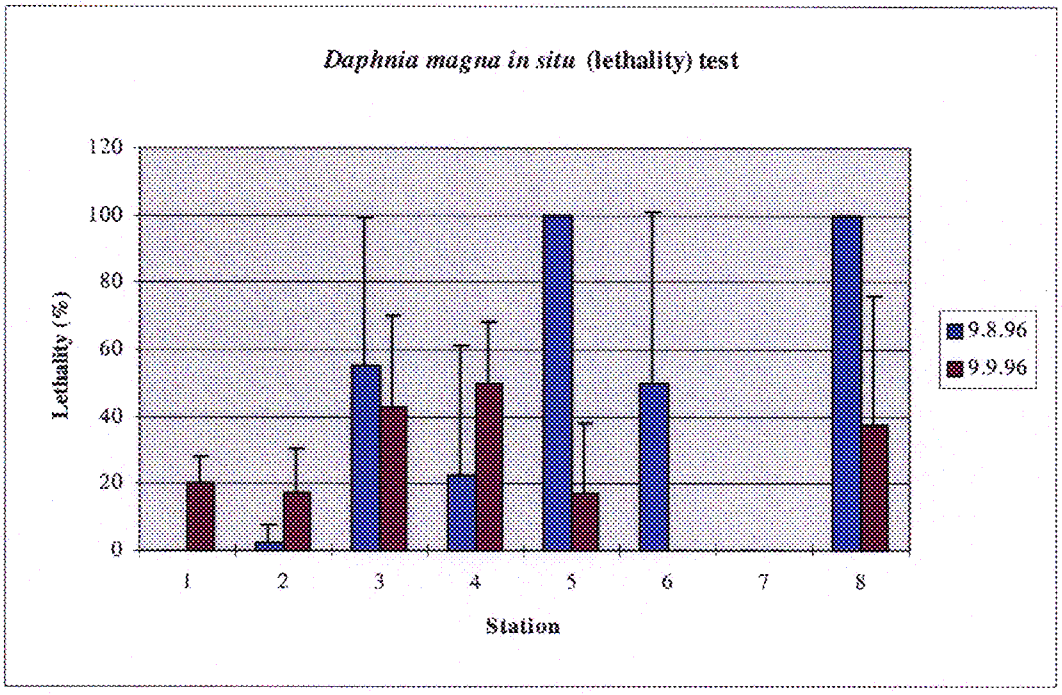

Figure 3.36 Mean lethality of *Daphnia magna* during two *in situ* deployments in the River Aire (error bars represent one standard deviation of cage replicates from the mean, ND = not deployed, * = no lethality measured, L = lost)

Table 3.31 shows the results of three and two way ANOVA and the ranking of responses between stations.

As seen for the laboratory bioassays with *Daphnia*, mean lethality in the *in situ* bioassays was different for the two deployments. The three way ANOVA showed this difference in deployment time to be statistically significant (that is, variance ratio = 8.44, $p<0.05$, df=1). This may have been due to differences in flow regime and discharge of contaminants during the exposure duration. Two way ANOVA demonstrated that on the first deployment there were statistically significant differences between stations. The test showed no discrimination on the second occasion (that is, no station effect).

Variability seen between cages as observed by the error bars in Figure 3.36 and three way ANOVA shows cage effects to be significant (see Table 3.31). This could be due to predation and/or a decline in water quality in some cages. Statistical power and discrimination could be improved by an increasing number of replicates and number of organisms deployed within each cage (see Section 3.4). For this study, four cages were deployed containing 10 animals each.

Table 3.31 Results of ANOVA for the *Daphnia magna in situ* bioassay results from the River Aire and ranking of responses

Site		Variance ratio and level of probability				Ranking of responses at stations (lowest to highest)
	Time	Time	Cage	Station	Time vs Station	
River Aire	2 & 3	8.44 p<0.05	5.36 p<0.01	11.1 p<0.001	14.6 p<0.001	1=2,4,3,5,8
	2	-	-	7.21 p<0.05	-	1,2,4,6,5,8
	3	-	-	1.39 NS	-	2=5,1,8,3,4

***Chironomus riparius in situ* (lethality) bioassay**

Figure 3.37 shows the results of *Chironomus riparius in situ* bioassays performed on two occasions for the River Aire. Cages were not deployed at Stations 6 and 7 for the same reasons no sediment samples could be taken from these stations. No deployment was possible at Station 2 on the second sampling occasion due to logistical constraints.

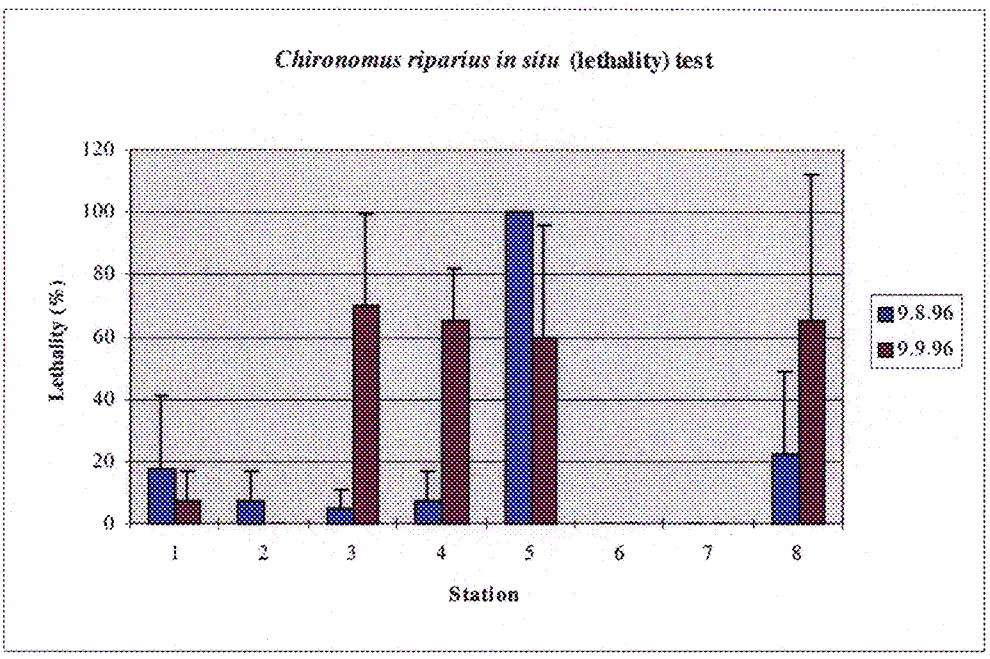

Figure 3.37 Mean *in situ* lethality of *Chironomus riparius* during two *in situ* deployments in the River Aire (error bars represent one standard deviation of cage replicates from the mean, ND - not deployed)

Table 3.32 shows the results of the three and two way ANOVA and the ranking of responses between stations.

Table 3.32 Results of ANOVA for the *Chironomus riparius in situ* bioassay data from the River Aire and ranking of responses

Site	Time	Variance ratio and level of probability				Ranking of responses at stations (lowest to highest)
		Time	Cage	Station	Time vs Station	
River Aire	2 & 3	9.95 p<0.01	3.5 NS	6.17 p<0.01	5.6 p<0.01	1,4,3,8,5
	2	-	-	20.3 p<0.001	-	3,2=4,1,8,5
	3	-	-	2.74 NS	-	1,5,4=8,3

As seen for the 10 day lethality and 28 day growth/lethality laboratory bioassays with *Chironomus*, mean lethality in the *in situ* deployments was different for the two deployment occasions. Three way ANOVA showed that this difference in deployment date was statistically significant (variance ratio = 9.95, $p<0.01$, df=1). As with the laboratory tests this may have been due to differences in flow regime and discharge of contaminants during the exposure duration. A lower level of lethality was observed at most stations in the first deployment than in the second. This is similar to the result of the 10 day *Chironomus* laboratory bioassay in which lethality was low in the spot-samples taken at the start of the *in situ* deployment. These samples were taken during a period of higher flow following a rain event. Although weather conditions at the River Aire were not monitored for the remainder of the deployment it is possible that this higher flow may have persisted for the duration of the *in situ* exposure thus diluting any discharge of toxicants over this time. However, even though <20% lethality was seen at the first four stations, 100% lethality was observed at Station 5 (the positive control). This is consistent with the *Gammarus* feeding rate measurements made over the same exposure period and with the results of the 10 day *Chironomus* laboratory bioassay and *Daphnia in situ* bioassay performed on the second sampling occasion. As discussed previously, this station was downstream of a sewage discharge and the sediment was noticeably anoxic and sulphurous. Whether the lethality observed in the *Chironomus in situ* bioassay was due to anoxia or toxicant concentrations was not possible to determine without measurement of toxicant concentrations.

On the first sampling occasion ANOVA demonstrated that there were statistically significant differences between stations. These results were not apparently correlated with the 10 day lethality and 28 day growth/lethality laboratory tests with *Chironomus*. On the second sampling occasion a higher level of lethality was seen at Stations 3,4, and 8 although a lower level was seen at Station 5. However, at this time ANOVA demonstrated that the differences between stations were not significant in determining the pattern of response. This may have been due to the high levels of variability within sites as demonstrated by the error bars in Figure 3.37. In some cases the contents of the cages were lost due to predation (leeches were often found inside the cages despite the small mesh size) and a decline in water quality within the cage due to blockage of the mesh with plant material and sediment. An increase in the statistical power of this test could be achieved by increasing the number of replicate cages and

the number of animals deployed within each cage (see Section 3.4). For this study, four cages were deployed containing 10 animals each.

Despite the lack of discrimination on the second sampling occasion, the results measured in this deployment were consistent with the results of the 10 day *Chironomus* laboratory test performed on samples taken at the beginning of the deployment. This observation was made despite the fact that the *Chironomus in situ* bioassays are not strictly measuring whole sediment toxicity. The test organisms in these chambers were only in contact with the sediment surface though a 243 µm mesh. However, fine sediment was found inside the cages on retrieval and this particle size fraction is likely to have contained high levels of toxicants if present in addition to any dissolved toxicants leaching from the sediment porewater. An alternative cage design was trialed during the course of this study in which animals were placed in tubes open to the sediment surface. However, in these cages predation was high, the test organisms were difficult to differentiate from indigenous organisms and sieving of the cage contents on retrieval was labour intensive.

3.1.6 *In situ* bioassays - estuarine

Mytilus edulis in situ (clearance rate) test

Figures 3.38 and 3.39 show the results of *M. edulis in situ* bioassay (with clearance rate as the endpoint) performed on two occasions for the Tees and Mersey estuaries.

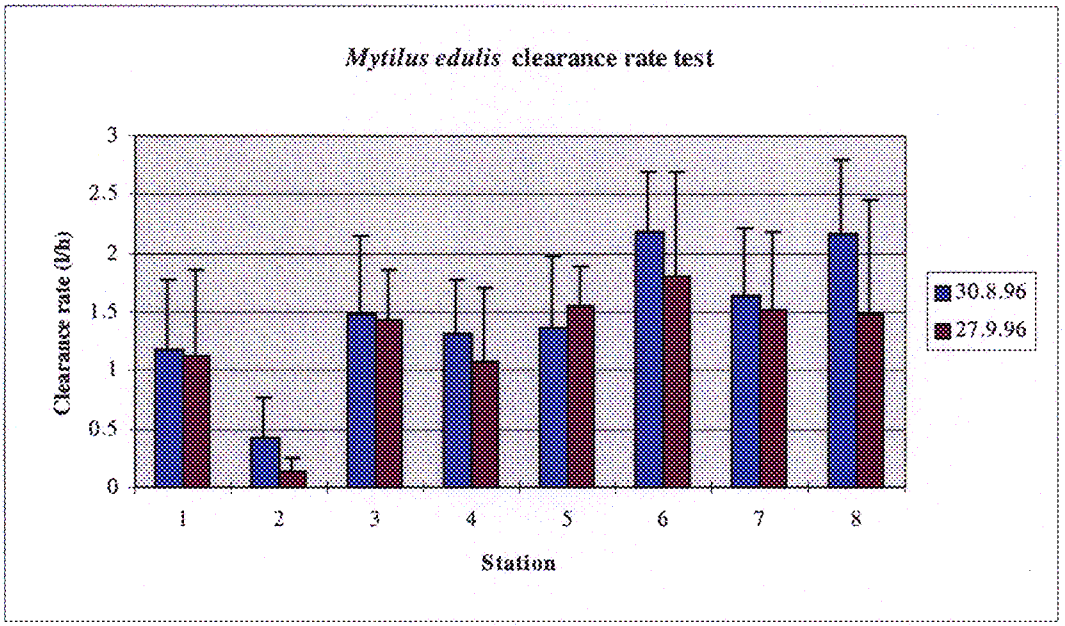

Figure 3.38 Mean clearance rate of *Mytilus edulis* during two *in situ* deployments in the Tees Estuary (error bars represent one standard deviation from the mean)

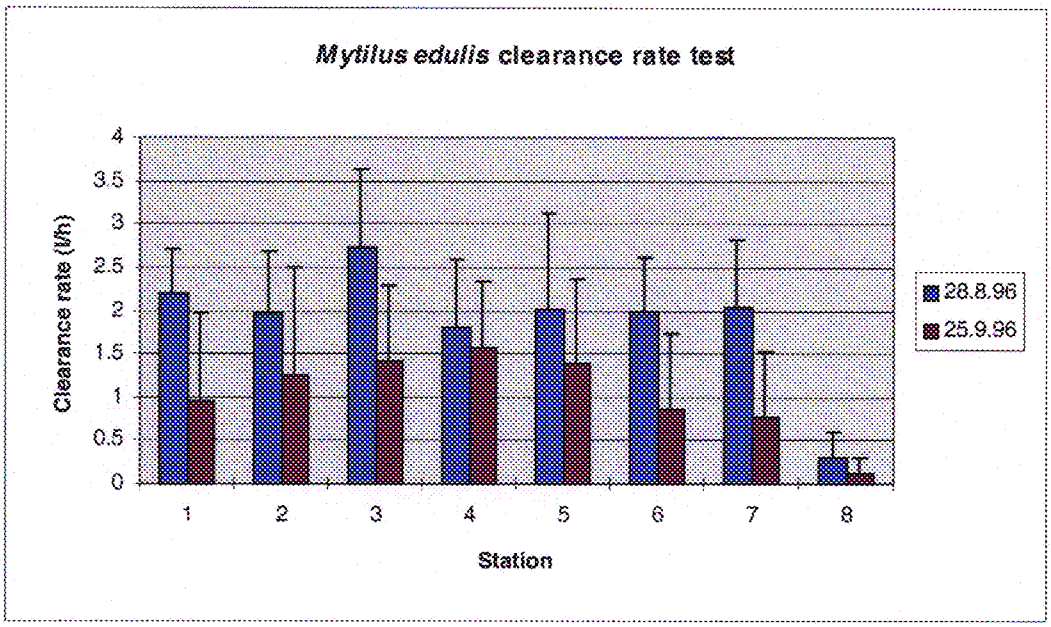

Figure 3.39 Mean clearance rate of *Mytilus edulis* during two *in situ* deployments in the Mersey Estuary (error bars represent one standard deviation from the mean)

Table 3.33 shows the results of the three and two way ANOVA and the ranking of responses between stations. ANOVA demonstrated that there were differences between stations at both the Tees and Mersey estuaries. This was the case when both deployment occasions were considered together and separately (see Table 3.33).

Table 3.33 Results of ANOVA for the *Mytilus edulis in situ* bioassay data from the Tees and Mersey estuaries and the ranking of responses

Site	Time	Variance ratio and level of probability				Ranking of responses at stations (highest to lowest)
		Time	Cage	Station	Time vs Station	
Tees Estuary	2 & 3	5.26 p<0.05	-	16.1 p<0.001	1.03 NS	6,8,7,5,3,4,1,2
	2	-	-	11.1 p<0.001	-	6,8,7,3,5,4,1,2
	3	-	-	6.7 p<0.001	-	6,5,7,8,3,1,4,2
Mersey Estuary	2 & 3	46.0 p<0.001	-	9.59 p<0.001	1.78 NS	3,5,4,2,1,6,7,8
	2	-	-	9.2 p<0.001	-	3,1,7,5,6,2,4,8
	3	-	-	3.16 p<0.01	-	4,3,5,2,1,6,7,8

The 2-way ANOVA showed the differences in mean *M. edulis* clearance rates with deployment date to be statistically significant for both sites (Tees Estuary: variance ratio = 5.26, $p<0.05$, df=7, Mersey Estuary: variance ratio = 46.0, $p<0.001$, df=7 respectively). This may be due to differences in the patterns of discharge of contaminants.

On the Tees Estuary, clearance rate was markedly lower at Station 2 compared to the other stations. Stations 6, 7 and 8 showed the highest clearance rates for each deployment. On the Mersey Estuary clearance rate was markedly lower at Station 8 compared to all other stations.

Corophium volutator in situ (lethality) bioassay

Figure 3.40 shows the results of *Corophium volutator* lethality *in situ* bioassays performed on two occasions for the Tees Estuary. Cages were not deployed at Station 6 as access was problematic.

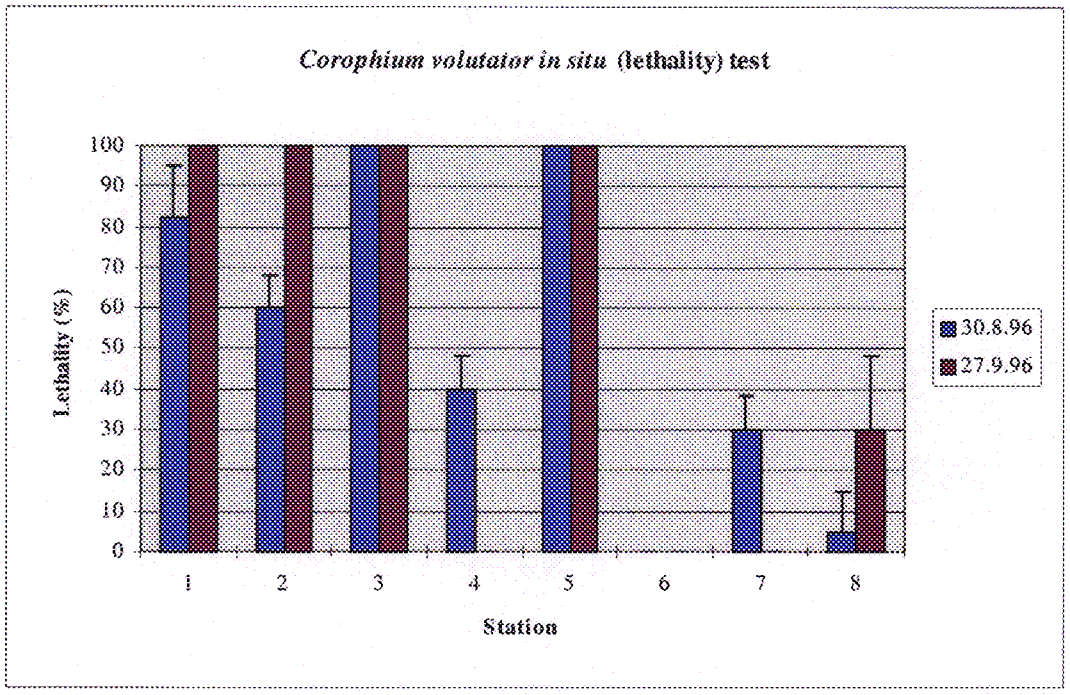

Figure 3.40 Mean lethality of *Corophium volutator* during two *in situ* deployments in the Tees estuary (error bars represent one standard deviation of cage replicates from the mean, ND = not deployed, L = lost).

Table 3.34 shows the results of the three and two way ANOVA and the ranking of responses between stations.

Table 3.34 Results of ANOVA for the *Corophium volutator in situ* bioassay data from the Tees Estuary and ranking of responses

Station	Time	Variance ratio and level of probability				Ranking of responses at stations (lowest to highest)
		Time	Cage	Station	Time vs Station	
Tees Estuary	2 & 3	293.8 p<0.001	5.78 p<0.05	327 p<0.001	36.4 p<0.001	8,1,2,3=5
	2	-	-	71.4 p<0.001	-	8,7,4,2,1,3=5
	3	-	-	11.3 p<0.001	-	8,1=2=3=5

As was evident for the 10 day lethality and 28 day growth/lethality laboratory bioassays with *Corophium*, mean lethality in the *in situ* bioassay was different for the two deployment occasions (see Table 3.34). The three way ANOVA showed this to be statistically significant (variance ratio = 293.8, $p<0.001$, df=1). However, on the first occasion the cages were retrieved after nine days due to the logistical constraints of site access. On the second occasion the cages were retrieved after 10 days which was the intended exposure duration selected to be comparable with the laboratory lethality bioassay. On the second occasion a high level of mortality was seen in all cages with the exception of those deployed at Station 8 which may reflect differences in the sensitivity of organisms between deployments. On the second sampling occasion the cages at Station 7 had been removed by tidal action and Station 4 cages could not be retrieved.

On the first sampling occasion, two way ANOVA demonstrated that differences between stations were significant in determining the pattern of *Corophium in situ* lethality. The pattern of results was consistent with the biological classification (see Table 1.6) and also with results of the 10 day lethality and 28 day growth/lethality laboratory bioassays which showed Stations 3 and 5 to be consistently grouped in the highest band of lethality, Station 4 in the intermediate band and Station 8 to be in the lowest band. A more consistent pattern was seen between the laboratory and *in situ* tests for *Corophium* than was seen for the *Chironomus* laboratory based sediment bioassays and *in situ* test. This could be because the *C.volutator in situ* bioassay was a whole sediment test. The exposure chambers were placed into the sediment and animals added directly to the sediment surface in which they were observed to burrow immediately. Whole sediment tests are obviously more desirable if the objective is to measure sediment toxicity. However, this is easier to achieve in the estuarine situation as the sediment is accessible at low tide. Also the chances of including organisms large enough to prey on *Corophium* within the exposure chambers are reduced when compared to the freshwater system.

The one potential disadvantage of placing *Corophium* directly into the sediment is that the bioassay response may potentially be influenced by sediment type as observed for some of the laboratory sediment tests. The three way ANOVA demonstrated that differences between cages were significant in determining patterns of response (that is, variance ratio = 5.78, $p<0.05$, df=3) although as demonstrated by the error bars in Figure 3.40 the within-site variability was considerably less than for the freshwater *in situ* tests. Differences between

cages in this test may have been due to differences in sediment type which may have also contributed to the patterns seen between stations. However, there were no significant correlations between the results of this test and the sediment factors measured. If these tests were selected, the influence of sediment type would have to be determined as discussed previously for the laboratory sediment tests.

3.1.7 Biomarkers

Metallothionein concentrations

Mussels from Stations 1, 4 and 7 on the Mersey had significantly elevated mean MT concentrations when compared with the other stations ($P<0.05$; Figure 3.41). Mean MT concentrations ranged from 86-114 $\mu g\, g^{-1}$ for these stations. The other stations (2,3,5,6 and 8) had mean MT concentrations ranging from 60-68 $\mu g\, g^{-1}$.

Figure 3.42 shows the mean levels of MT in mussels transplanted in the Tees Estuary. There were no significant differences in mean MT concentrations among the stations ($P>0.05$), although all mean MT levels were all >86.6 $\mu g\, g^{-1}$.

MT concentrations in mussels deployed at Stations 2, 3, 5 and 6 on the Mersey estuary were within the range found for other mussel populations obtained from clean sites (up to 80 $\mu g\, g^{-1}$; Cheung, unpublished). Mussels from the remaining sites on the Mersey, and from all the sites on the Tees, exhibited elevated levels of MT (>86.6 $\mu g\, g^{-1}$). The elevated concentrations of MT present at some of the stations could indicate an adaptive response to trace metal contamination. However, in future it is necessary to measure metal concentrations and relative proportion of metals bound to MT in order to gain a more integrated view.

Lysosome integrity

Overall, there were no significant differences among the mean retention times for any of the stations on the Mersey ($P>0.05$; Figure 3.43). However, multiple range tests indicated that mussels from Station 5 had significantly higher mean retention times (59 min) than Stations 6, 7 and 8 (ranging from 38-42 min; $P<0.05$).

Stations 1, 2 and 8 on the Tees had significantly higher mean retention times compared with the remaining sites ($P<0.05$; Figure 3.44). There were no significant differences between Stations 4, 5, 6 and 7 ($P>0,.05$). Stations 5 and 6 had the lowest mean retention times (41 and 39 min, respectively).

Commonly, in an uncontaminated site, neutral red retention values for molluscan haemocytes have been recorded as being in the range 100-150 min (Wedderburn, unpublished). All lysosome retention times were <100 min on both the estuaries. Lysosomal retention was shown to be substantially depressed at Stations 4, 5 and 6 on the Tees estuary (<55 minutes). The mean retention values for the Mersey (44 min) were lower than the Tees (66 min) and there was no significant difference between the individual stations. The low values in both estuaries recorded are in accordance with values obtained in other studies from contaminated coastal sites. Lowe *et al.* (1993) found that the retention of neutral red was significantly reduced (<30 min) in haemocytic lysosomes in mussels collected from severely contaminated

sites in the Venice lagoon. Moncheva (pers. comm.) recorded retention times in the range of 15-49 minutes for an impacted harbour site on the Black Sea coastline. Wedderburn (unpublished) recorded values in the range 60-70 minutes for a moderately contaminated site in southwest England.

Heart rates

Heart rates of mussels were found to be irregular over the first hour of cardiac recording, which was invariably due to handling stress caused by transducer attachment. Thus, the first hour of recordings were not included.

Box and Whisper plots for the groups of mussels from the Mersey indicate that median heart rates were significantly lower at Station 8 ($P<0.05$; Figure 3.45). Heart rates ranged from 2-6 beats per minute at this station. Median heart rates were not significantly different among the other stations ($P>0.05$). The majority of the mussels at these stations had median heart rates ranging between 20-28 beats per minute.

For the groups of mussels deployed in the Tees, there were significant differences in median heart rates among the different stations ($P<0.05$). Mussels from Stations 1, 4 and 5 had significantly lower median heart rates than Stations 2, 7 and 8 ($P<0.05$; Figure 3.46; there was large inter-individual variability at Station 6). Nevertheless, with the exception of Station 5, the majority of mussels had median heart rates ranging between 20-28 beats per minute. At Station 5, four out of the eight mussels had median heart rates lower than 15 beats per minute.

The non-invasive cardiac monitoring system used in this study proved to be reliable for recording heart rates from groups of 16 mussels, simultaneously, for 7 h. There is no reason why recordings could not have been continued for several days or weeks. The absolute heart rates reported here are within the range found for other populations of mussels occupying estuarine habitats (mussels 4-5 cm length, 20-28 beats per minute at 15 °C; Bayne *et al.*, 1976). The reduced heart rates observed at Station 8 on the Mersey and Station 5 on the Tees are typical of bivalve molluscs in poor condition (Coleman, 1974). The fall in heart rate at these sites are consistent with the results of the estuarine bioassays and would appear to be indicative of pollutant toxicity. For instance, Grace and Gainey (1987) reported decreased heart rates for *Mytilus edulis* after four day's exposure to five concentrations of copper (between 0.05-0.4 ppm). In addition, Sabouarin and Tullis (1981) demonstrated that high concentrations of three aromatic hydrocarbons caused significant decreases in heart activity for *Mytilus californianus.*

The stations shown to have elevated levels of MT do not correlate with those which the neutral red and cardiac techniques indicate as being impacted, indicating a mixture of substances (trace metals and organic compounds) may be present in the receiving waters. This may be due to the temporal difference between the deployment of the mussels used for MT analysis and those transplanted for neutral red and heart rate. On the Tees, the neutral red and cardiac data suggest that Station 5 is of lower environmental quality than the other stations on this estuary. This is in agreement with the *T. battagliai* lethality and OEL tests carried out on water samples and *C. volutator* tests on sediment samples. In conclusion, these biomarker techniques have identified poor environmental quality at the majority of stations on both estuaries, and have highlighted two specific stations where conditions were particularly poor.

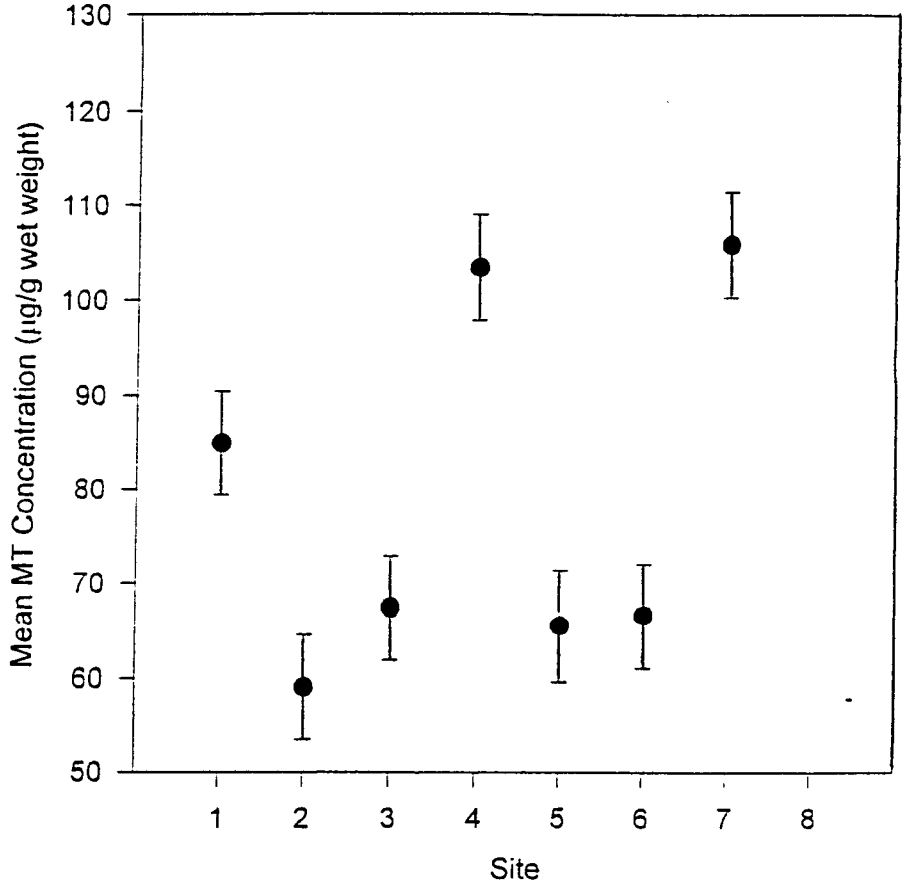

Figure 3.41 Mean levels of MT (and 95% confidence levels) of mussels from seven sites along the Mersey Estuary (Station 8 missing)

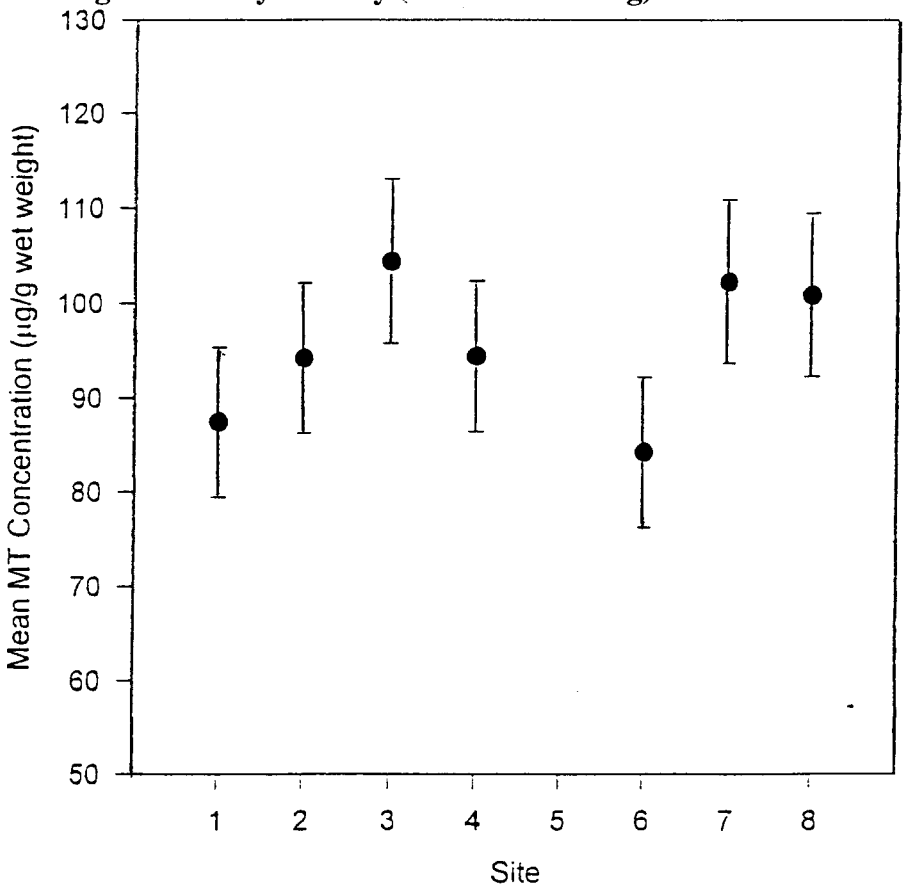

Figure 3.42 Mean levels of MT (and 95% confidence levels) of mussels from seven sites along the Tees Estuary (Station 5 missing)

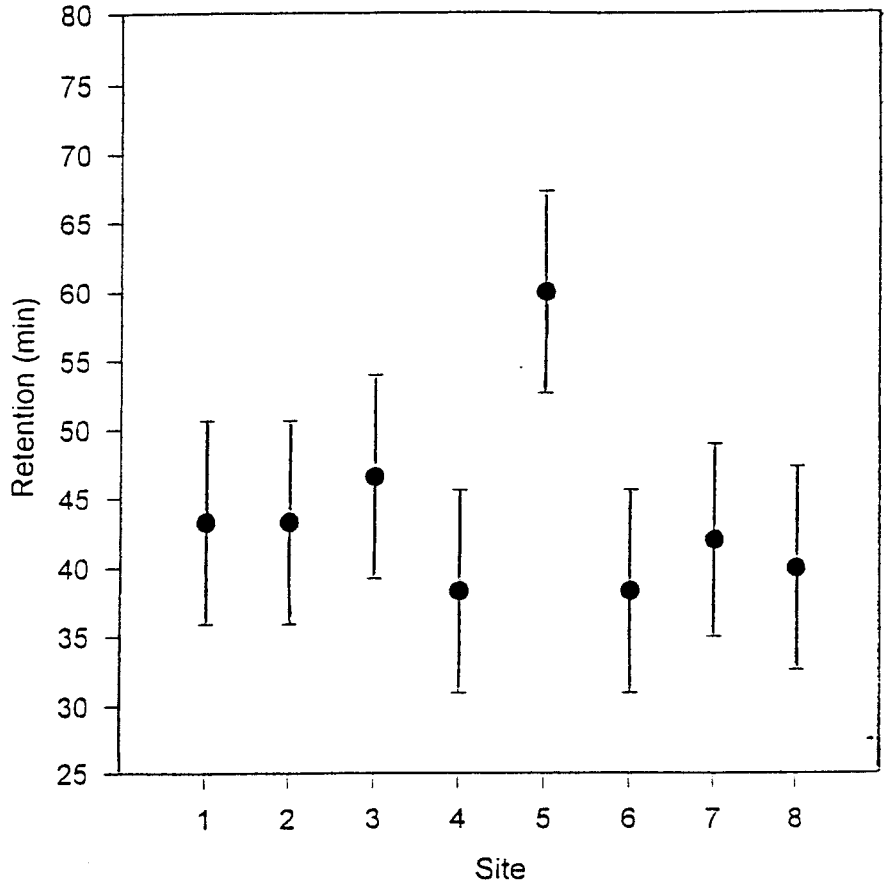

Figure 3.43 Mean neutral red retention times (and 95% confidence levels) of mussels from eight sites along the Mersey Estuary

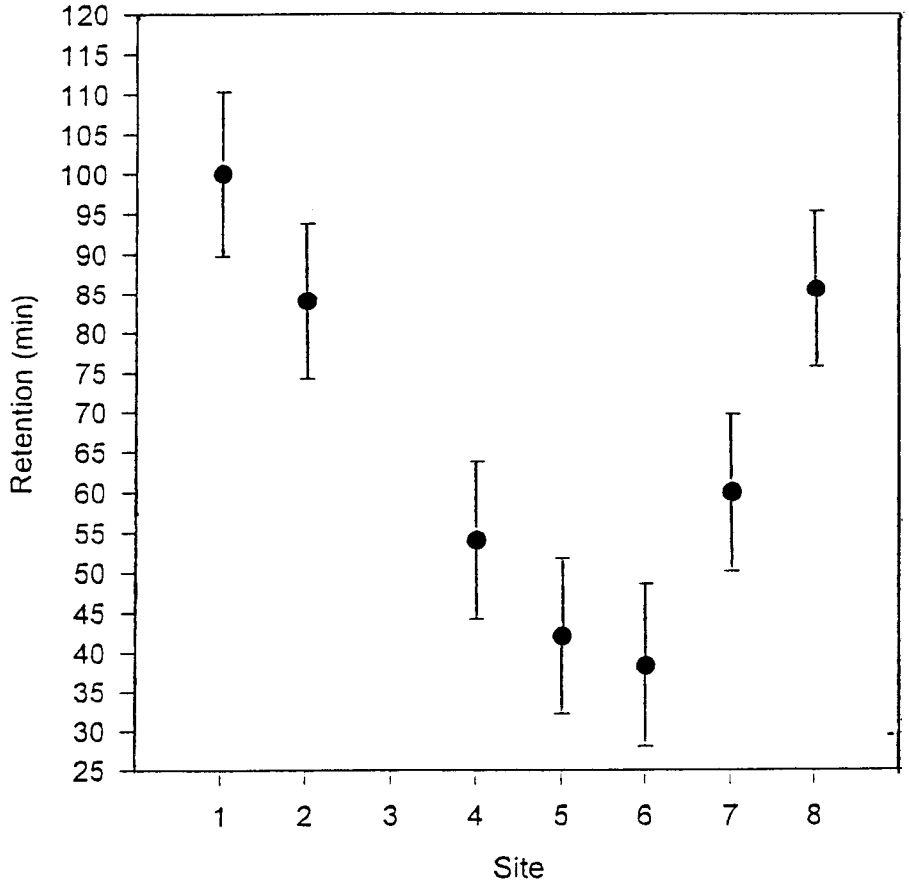

Figure 3.44 Mean neutral red retention times (and 95% confidence levels) of mussels from seven sites along the Tees Estuary (Station 3 missing)

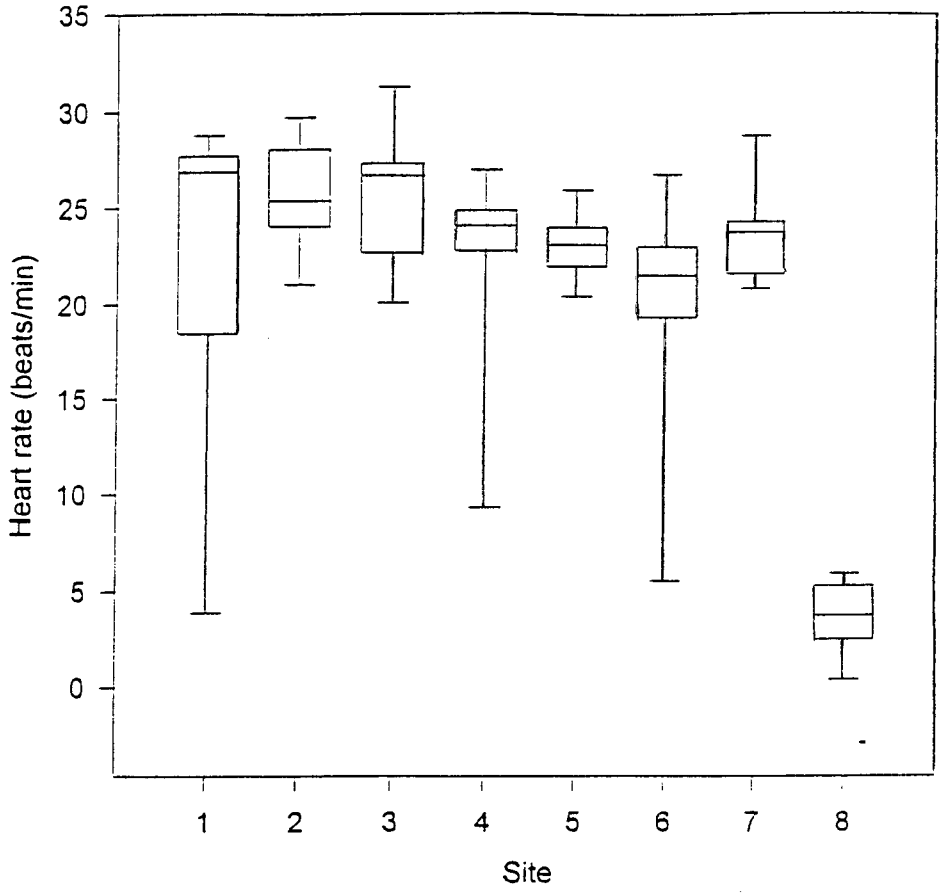

Figure 3.45 Box and Whisper plots for median heart rate of mussels from eight sites along the Mersey Estuary

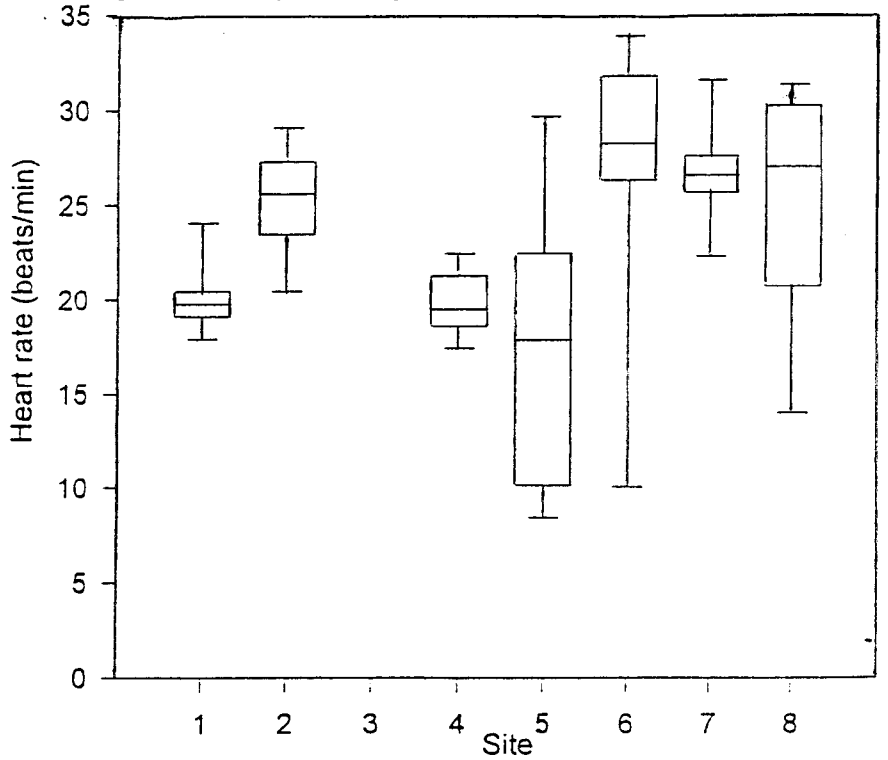

Figure 3.46 Box and Whisper plots for median heart rate of mussels from seven sites along the Tees Estuary (Station 3 missing)

3.2 Correlations between bioassays

On the basis that any general quality assessments or local environmental impact assessments carried out using bioassays will need to use a test battery the data from tests of the same types (for example laboratory based water column) used at each site has been compared to quantify the nature of correlations. If two bioassays of a particular type respond in the same manner a range of situations then there would redundancy if both were used in the test battery.

Table 3.35 summarises the correlations between the different laboratory-based freshwater bioassays. At the River Aire the ECL bioassay responses were correlated with 10 d *D. magna* lethality data. No statistically significant correlations were found between ECL bioassay responses and *L. minor* growth inhibition responses at both sites, *D. magna* juvenile production data for the River Aire or *D. magna* juvenile production or lethality data for the Willow Brook. There were no statistically significant correlations between Microtox-acute data and that for the *L. minor* growth inhibition responses and *D. magna* reproduction bioassay endpoints at both sites. At the River Aire *L. minor* growth inhibition responses were correlated with those of *D. magna* juvenile production. The two responses in the *D. magna* reproduction bioassay (juvenile production and lethality) showed a significant correlation for the River Aire samples.

Table 3.36 summarises the correlations between the different laboratory-based estuarine water column bioassays. There were statistically significant correlations between a number of bioassays. ECL bioassay responses and the OEL PNR values were correlated at both sites. At the Tees Estuary ECL bioassay responses and *T. battagliai* juvenile production were significantly correlated while at the Mersey Estuary ECL responses and *T. battagliai* lethality were significantly correlated. Microtox-acute bioassay responses were significantly correlated with OEL and *T. battagliai* reproduction bioassay endpoint data at both sites. The OEL test data were significantly correlated with the *T. battagliai* lethality data for both sites and with juvenile production data for the Tees Estuary site. The responses in the *T. battagliai* reproduction bioassay (juvenile production and lethality) were significantly correlated at both sites.

Table 3.37 summarises the correlations between the different laboratory-based freshwater sediment bioassays. The responses of the 10 d *C. riparius* lethality bioassay were not significantly correlated with any of the endpoints in the 28 d *C. riparius* growth/lethality bioassay. The growth responses in the 28 d bioassay were significantly correlated with lethality but not development responses and the development and lethality responses were significantly correlated.

Table 3.38 summarises the correlations between the different laboratory-based estuarine sediment bioassays. The responses of the 10 d *C. volutator* lethality bioassay were not significantly correlated with the growth or lethality endpoints of the 28 d *C. volutator* bioassay at either site. At the Tees Estuary site the only significant correlation between the 10 d and 28 d *C. volutator* bioassays and the 10 d *A. marina* bioassay was between the lethality response in the 28 d *C. volutator* bioassay and cast formation in the 10 d *A marina* bioassay.

The only significant correlation between the responses of freshwater *in situ* bioassays deployed at the River Aire was the negative correlation between lethality in the *Daphnia magna in situ* bioassay and feeding rate in the *Gammarus pulex in situ* bioassay($r=-0.79$, $p<0.01$, $n=12$).

nses of different laboratory based freshwater water column

	on coefficient (and significance) between bioassays		
	L. minor growth inhibition (7 d)	*D. magna* reproduction (10 d) Juvenile production	Lethality
	0.407 (NS)	0.017 (NS)	**0.488(p<0.01)**
	0.006 (NS)	-0.031 (NS)	-0.102 (NS)
	-	0.409 (NS)	-0.302 (NS)
		-	**-0.621 (p<0.001)**
			-
5)	-0.141 (NS)	0.122 (NS)	0.254 (NS)
	-0.402 (NS)	0.150 (NS)	-0.08 (NS)
	-	**-0.605 (p<0.05)**	0.138 (NS)
		-	-0.178 (NS)
			-

Table 3.36 Correlation coefficients (and significance) between responses of different laboratory based estuarine water column bioassays

Site	Bioassay	Correlation coefficient (and significance) between bioassays				
		ECL	Microtox-acute	OEL	*T. battagliai* reproduction (9 d) Juvenile production	Lethality
Tees Estuary	ECL	-	-0.095 (NS)	**0.485 (p<0.01)**	**-0.363 (p<0.05)**	0.127 (NS)
	Microtox-acute		-	**0.735 (p<0.001)**	**0.602 (p<0.001)**	**-0.735 (p<0.001)**
	OEL			-	**-0.615 (p<0.001)**	**0.81 (p<0.001)**
	T. battagliai reproduction (9 d)					
	- juvenile production				-	**-0.47 (p<0.01)**
	- lethality					-
Mersey Estuary	ECL	-	-0.4 (NS)	**0.961 (p<0.001)**	-0.396 (NS)	**0.831 (p<0.001)**
	Microtox-acute		-	**-0.409 (p<0.05)**	**0.76 (p<0.001)**	**-0.439 (p<0.05)**
	OEL			-	-0.395 (NS)	**0.835 (p<0.001)**
	T. battagliai reproduction (9 d)					
	- juvenile production				-	**-0.5 (p<0.05)**
	- lethality					-

ificance) between responses of different laboratory based freshwater sediment bioassays

Correlation coefficient (and significance) between bioassays			
C. riparius ethality (10 d)	*C. riparius* growth (28 d)	*C. riparius* development (28 d)	*C. riparius* lethality (28 d)
-	-0.162 (NS)	-0.179 (NS)	-0.01 NS)
	-	-0.351 (NS)	**0.513 (p <0.01)**
		-	**-0.704 (p<0.001)**

Table 3.39 Summary of the data for freshwater bioassays used at reference Station 1 and positive control Station 5 on the River Aire

Test used (endpoint)	Responses (SD) at different stations on each sampling occasion												Achieving validity criterion
	Reference station				*Positive control station*				Difference (Significance)				
	T1	T2	T3	T4	*T1*	*T2*	*T3*	*T4*	T1	T2	T3	T4	
ECL (ECL Units)	10.3 (0.5)	5.5 (1.7)	4.1 (1.9)	2.1 (1.9)	*80.5 (14)*	*32.2 (15)*	*45.9 (4.4)*	*48.6 (13)*	70.2	26.7	41.8	46.7	NV
Microtox acute (% of control)	120.9 (5.7)	100.7 (6.8)	118 (15)	112.4 (6.3)	*126.7 (4.2)*	*119.7 (3.7)*	*100.8 (38)*	*122.6 (6.1)*	5.8	19.0	-17.2	10.2	NV
7 d *L. minor* growth (growth rate)	0.270 (0.03)	0.273 (0.07)	NT	NT	*0.285 (0.04)*	*0.298 (0.07)*	*NT*	*NT*	0.015	0.025	NT	NT	NV
10 d *D.magna* reproduction (% lethality)	4	4.2	0	4	*84*	*0*	*100*	*96*	80	0	100	92	⌋,⌋,⌋,⌋
10 d *D.magna* reproduction (No. young)	NC	14.7 (6.6)	NC	NC	*NC*	*14.6 (4.5)*	*NC*	*NC*	NC	0.1	NC	NC	NV
10 d *C. riparius* lethality (% lethality)	NT	25.6 (15)	53.0 (8.2)	NT	*NT*	*11.5 (6.4)*	*90.2 (6.7)*	*NT*	NT	-14.1	37.2	NT	x, x
28 d *C. riparius* growth (% lethality)	NT	30.7 (11)	TF	NT	*NT*	*49.3 (12)*	*TF*	*NT*	NT	18.6	TF	NT	NV
28 d *C. riparius* growth (% development)	NT	46.1 (18)	TF	NT	*NT*	*37 (9)*	*TF*	*NT*	NT	-9.1	TF	NT	NV
28 d *C. riparius* growth (Growth)	NT	0.791 (0.17)	TF	NT	*NT*	*0.923 (0.36)*	*TF*	*NT*	NT	0.132	TF	NT	NV
6 d *G. pulex in situ* (% lethality)	NT	18	2	NT	*NT*	*48*	*36*	*NT*	NT	30	34	NT	NV
6 d *G. pulex in situ* (Feeding rate)	NT	0.226 (0.04)	0.429 (0.01)	NT	*NT*	*0 (0.05)*	*0.138 (0.03)*	*NT*	NT	0.226	0.291	NT	NV
6 d *D.magna in situ* (% lethality)	NT	0 (0)	20 (8.2)	NT	*NT*	*100 (0)*	*17.5 (21)*	*NT*	NT	100	-2.5	NT	NV
10 d *C. riparius in situ* (% lethality)	NT	17.5 (24)	7.5 (10)	NT	*NT*	*100 (0)*	*60 (36)*	*NT*	NT	82.5	52.5	NT	NV

Key: T1, T2, T3, T4: Sampling/deployment times 1 (10/7/96), 2 (9/8/96), 3 (9/9/96) and 4 (17/10/96), NT: Not tested, TF: Test failed, NC: Not calculated, See Section 2.2.3 for information on validity criteria, ⌋ = criterion achieved on a particular occasion, x = validity criterion not achieved, NV: No appropriate test validity criterion

Table 3.40 Summary of the statistical analysis of the freshwater bioassay data for Stations 1 and 5 on the River Aire

Tests used (and endpoints)	Statistical procedure (test statistic)	Test statistic value (significance)			Comments
		Station	Time	Station x Time	
ECL (ECL Units)	Kruskal Wallis (H)	100.3 (p<0.001)	7.1 (NS)	12.0 (p<0.01)	Significant increases in response at Station 5, relative to Station 1 on all occasions, though the increase was lower at time 2
Microtox acute (% of control)	Kruskal Wallis (H)	53.6 (p<0.001)	6.2 (NS)	2.8 (NS)	Significant increases in response at Station 5, relative to Station 1, but no consistent pattern over time
7 d *L. minor* growth (Growth rate)	Kruskal Wallis (H)	3.2 (NS)	1.7 (NS)	14.7 (p<0.01)	No significant decreases in growth at Station 5, relative to Station 1
10 d *D.magna* reproduction (% lethality)	Kruskal Wallis (H)	76.9 (p<0.001)	8.8 (p<0.05)	3.7 (NS)	Significant increase in lethality at Station 5, relative to low levels at Station 1, with effects being time dependent and only evident under low flow conditions (times 1,3 and 4).
10 d *D.magna* reproduction (No. young)	Data not analysed because of complete adult mortality at positive control station				
10 d *C. riparius* lethality (% lethality)	Kruskal Wallis (H)	0.12 (NS)	11.4 (p<0.01)	18.6 (p<0.001)	No significant increases in lethality at Station 5 relative to Station 1
28 d *C. riparius* growth (% lethality)	Mann-Whitney (U)	25 (p<0.01)	-	-	Significant increases in lethality at Station 5 relative to Station 1
28 d *C. riparius* growth (% development)	Mann-Whitney (U)	5.5 (NS)	-	-	No significant difference in development at Station 5 relative to Station 1
28 d *C. riparius* growth (Growth)	Mann-Whitney (U)	8 (NS)	-	-	No significant difference in growth at Station 5 relative to Station 1
6 d *G. pulex in situ* (% lethality)	Kruskal Wallis (H)	4.2 (p<0.005)	0.93 (NS)	6.1 (p<0.05)	Significant increases in lethality at Station 5 relative to Station 1
6 d *G. pulex in situ* (Feeding rate)	Kruskal Wallis (H)	4.2 (p<0.05)	0.82 (NS)	6.1 (p<0.05)	Significant decreases in feeding rate at Station 5 relative to Station 1
6 d *D.magna in situ* (% lethality)	Kruskal Wallis (H)	2.8 (NS)	0.12 (NS)	4.4 (p<0.05)	No significant increases in lethality at Station 5 relative to Station 1
10 d *C. riparius in situ* (% lethality)	Kruskal Wallis (H)	6.6 (p<0.05)	0.53 (NS)	0.09 (NS)	Significant increases in lethality at Station 5 relative to Station 1

In the 7 d *L. minor* growth inhibition bioassay there were small increases in growth rate (0.015-0.025 Units) in the Station 5 samples, relative to those from Station 1, on the two occasions they were tested, but these were not statistically significant ($H=3.2$, $p>0.05$, $df=1$) The increases in growth at Station 5 were indicative of eutrophication rather than toxicant-induced effects.

In the 10 d *D.magna* reproduction bioassay there was marked temporal variability in the level of lethality (after 10 days) in animals exposed to water samples from Stations 1 and 5, with the differences being significant ($H=76.9$, $p<0.001$, $df=1$) The lethality of *D. magna* at Station 5 (>84%) was markedly higher on three of the four occasions (times 1, 3 and 4), relative to the low levels found in Station 1 samples (<8% in all cases). On the second sampling occasion lethality was low in samples from both stations (<5%). Juvenile production was not different in the surviving adults at each station on the second sampling occasion. The level of lethality in adult daphnids exposed to the reference Station 1 water samples was less than the 20% validity criterion specified in the OECD guidelines for the 21 day *D.magna* juvenile production test on all sampling occasions (OECD 1998).

The responses in the 10 d *C. riparius* lethality bioassay at the two River Aire stations were not significantly different ($H=0.12$, $P>0.05$, $df=1$) but showed marked temporal variability ($H=11.4$, $p<0.001$, $df=3$). Lethality was similar at the positive control Station 5 to that at the reference location Station 1 on the first test occasion (time 2) and higher at Station 5 than Station 1 on the second test occasion (time 3). In the 10 d *Chironomus* sediment bioassay the level of lethality in the samples from the reference location only approximated the 20% test validity criterion (ASTM 1997) on the first sampling occasion. The high level of lethality measured in sediment bioassays of samples from the reference location on the second occasion (53%) was unexpected given the low level of lethality at downstream Station 2 (16%) and the low level of lethality at Station 1 on the first occasion. However, Station 1 was surrounded by farmland and it is possible that run-off of hydrophobic contaminants such as pesticides had occurred during the rainfall, or that an aqueous discharge from the farm had entered the system. If the increase in toxicity measured in the sediment is indeed due to chemical contaminants this indicates the usefulness of a direct toxicity assessment (DTA) approach in identifying a contamination episode which probably would not have been detected by chemical analysis of spot water samples. Although this impact may not translate into measurable effects on macroinvertebrate diversity and abundance a risk of impact has been identified and appropriate management action would need to be taken.

The 28 d *C. riparius* growth test was only successful on the first occasion it was used (time 2) and showed higher lethality in organisms exposed to Station 5 sediment compared to Station 1 sediment ($U=25$, $p<0.01$). However, there were no statistically significant differences between stations.

The 6 d *G. pulex* feeding rate bioassay showed significant differences in both the lethality ($H=4.2$, $p<0.05$, $df=1$) and feeding rate ($H=4.2$, $p<0.05$, $df=1$) endpoints between Station 1 (lethality 6-23.5%, feeding rate 0.249-0.429 mg mg^{-1} d^{-1}) and Station 5 (lethality 38-49%, feeding rate 0.06-0.149 mg mg^{-1} d^{-1}).

The 6 d *D.magna* lethality *in situ* bioassay did not show significant differences in lethality between stations ($H=2.8$, $p>0.05$, $df=1$). Large increases in lethality at Station 5 were

measured on the first deployment occasion (time 2) but no differences in effects were evident between the two stations on the second deployment occasion (time 3).

In contrast, the 10 d *C. riparius* lethality *in situ* bioassay showed significantly ($H=6.6$, $p<0.05$, $df=1$) higher lethality at Station 5 than Station 1 on both deployment occasions.

Estuarine bioassays

Table 3.41 shows the response data for each estuarine bioassay on each occasion they were deployed at the reference and positive control stations of the Tees Estuary along with information on the extent to which any test validity criteria were achieved. The results of statistical analyses (Kruskal-Wallis and Mann-Whitney tests) of the data are shown in Table 3.42. The statistical analyses used are described in detail in Section 2.3.

There were statistically significant differences in the effects measured in the samples from Station 5 compared to those from Station 8 in all laboratory water-column bioassays, that is the ECL test ($H=99.5$, $p<0.001$, $df=1$), the Microtox-acute test ($H=100$, $p<0.001$, $df=1$), the 24 h oyster embryo-larval development test ($H=100$, $p<0.001$, $df=1$) and the lethality endpoint of the 9 d *T. battagliai* reproduction test ($H=100$, $p<0.001$, $df=1$) on all sampling occasions (Stats). In both the OEL and *T. battagliai* tests there was complete abnormality and lethality of test organisms in samples from the positive control Station 5 on all occasions, whereas effects at the reference location Station 8 were in most instances below 10%. The levels of abnormalities in seawater controls of the OEL test were always less than the test validity criterion of 40% (ICES 1991).

The responses in the 10 d *C. volutator* lethality bioassay at the two Tees Estuary stations were significantly higher ($H=14.3$, $p<0.001$, $df=1$) at the positive control Station 5 than the reference location Station 8 on both test occasions (times 2 and 3).

Data was only obtained for the 28 d *C. volutator* growth test on the first occasion it was used (time 2) and this showed higher lethality in organisms exposed to Station 5 sediment compared to Station 8 sediment ($U=25$, $p<0.01$). No growth data was available for animals exposed to Station 5 sediment due to the complete lethality of all test organisms.

In the 10 d *A. marina* cast formation test there were no significant differences in lethality between Stations 5 (3.3-20.9%) and 8 (13.3-16.7%) on the two test occasions ($H=0.34$, $p<0.05$, $df=1$) but significant reductions in cast formation at Station 5 compared to Station 8 ($H=8.8$, $p<0.01$, $df=1$)

The 28 d *M. edulis* clearance rate *in situ* bioassay showed a significant station effect ($H=7.4$, $p<0.01$, $df=1$). There was a large reduction in clearance rate at Station 5, compared to Station 8, on the first deployment occasion (time 2) but no differences in effects at the two stations on the second deployment occasion (time 3).

In contrast, the *C. volutator* lethality *in situ* bioassay showed significantly ($H=7.5$, $p<0.01$, $df=1$) higher lethality at Station 5 (100% on both occasions) than Station 8 (5-30%) on the two deployment occasions.

Table 3.41 Summary of the data for estuarine bioassays used at reference Station 8 and positive control Station 5 on the Tees Estuary

Test used (endpoint)	Responses (SD) at different stations on each sampling occasion												Achieving validity criterion
	Reference station				*Positive control station*				Difference (Significance)				
	T1	T2	T3	T4	*T1*	*T2*	*T3*	*T4*	T1	T2	T3	T4	
ECL (ECL Units)	3.9 (0.7)	5.6 (2.6)	3.2 (0.5)	3.8 (0.9)	*250 (98)*	*53.7 (16)*	*36.2 (12)*	*96.3 (4.7)*	246	48.1	33.0	92.5	NV
Microtox acute (% of control)	79.4 (10)	67.0 (2.4)	60.4 (1.6)	66.9 (1.9)	*11.2 (1.4)*	*23.8 (6.1)*	*16.3 (2.4)*	*33.8 (3.1)*	68.2	43.2	44.1	33.1	NV
24 h OEL (Percent net response)	-2.0 (3.5)	-3.1 (3.0)	-1.2 (3.2)	6.6 (5.3)	*100 (0)*	*100 (0)*	*98 (1.7)*	*100 (0)*	102	103	99.2	93.4	⌋,⌋,⌋,⌋
9 d *T. battagliai* reproduction (% lethality)	8.3	4	0	30	*100*	*100*	*100*	*100*	91.7	96	100	70	⌋,⌋,⌋,x
9 d *T. battagliai* reproduction (No. young)	NC	NC	NC	NC	*NC*	*NC*	*NC*	*NC*	NC	NC	NC	NC	NV
10 d *C. volutator* lethality (% lethality)	NT	8.9 (6.1)	23 (16)	NT	*NT*	*99.6 (1.0)*	*94.3 (4.9)*	*NT*	NT	90.7	71.3	NT	⌋,x
28 d *C. volutator* growth (% lethality)	NT	20 (14)	TF	NT	*NT*	*100 (0)*	*TF*	*NT*	NT	80	TF	NT	NV
28 d *C. volutator* growth (Growth)	NT	NC	TF	NT	*NT*	*NC*	*TF*	*NT*	NT	NC	TF	NT	NV
10 d *A. marina* cast formation (% lethality)	NT	3.3	20.9	NT	*NT*	*13.3*	*16.7*	*NT*	NT	+10	-4.2	NT	NV
10 d *A. marina* cast formation (No. casts)	NT	5.3 (2.7)	6.1 (4.8)	NT	*NT*	*1.1*	*0*	*NT*	NT	4.2	6.1	NT	NV
6 d *M.edulis in situ* (Clearance rate)	NT	2.16 (0.63)	1.48 (0.98)	NT	*NT*	*1.37 (0.61)*	*1.55 (0.34)*	*NT*	NT	0.79	-0.07	NT	NV
10 d *C. volutator in situ* (% lethality)	NT	5 (10)	30 (18)	NT	*NT*	*100 (0)*	*100 (0)*	*NT*	NT	95	70	NT	NV

Key: T1, T2, T3, T4: Sampling/deployment times 1 (2/8/96), 2 (30/8/96), 3 (27/9/96) and 4 (24/10/96), NT: Not tested, TF: Test failed, NC: Not calculated,
See Section 2.2.3 for information on validity criteria, ⌋ = criterion achieved on a particular occasion, x = validity criterion not achieved, NV: No appropriate test validity criterion

Table 3.42 Summary of the statistical analysis of the estuarine bioassay data for Stations 8 and 5 on the Tees Estuary

Tests used (and endpoints)	Statistical procedure (test statistic)	Test statistic value (significance) Station	Time	Station x Time	Comments
ECL (ECL Units)	Kruskal Wallis (H)	99.5 (p<0.001)	5.1 (NS)	9.3 (p<0.05)	Significant increases in response at Station 5, relative to Station 8, on all occasions, though the increase was lower at time 3
Microtox acute (% of control)	Kruskal Wallis (H)	100 (p<0.001)	8.8 (p<0.05)	12.5 (p<0.01)	Significant decreases in response at Station 5, relative to Station 8, on all occasions
24 h OEL (Percent Net Response	Kruskal Wallis (H)	100 (p<0.001)	3.8 (NS)	10.6 (p<0.05)	Significant increases in response at Station 5, relative to Station 8, on all occasions
9 d *T. battagliai* reproduction (% lethality)	Kruskal Wallis (H)	100 (p<0.001)	0.94 (NS)	14.5 (p<0.01)	Significant increases in lethality at Station 5, relative to low levels at Station 8, on all occasions
9 d *T. battagliai* reproduction (No. young)	Data not analysed because of complete adult mortality at positive control Station 5				
10 d *C. volutator* lethality (% lethality)	Kruskal Wallis (H)	14.3 (p<0.001)	4.4 (p<0.05)	15.3 (p<0.001)	Significant increases in lethality at Station 5 relative to Station 8
28 d *C. volutator* growth (% lethality)	Mann-Whitney (U)	25 (p<0.05)	-	-	Significant increase in lethality at Station 5 relative to Station 8
28 d *C. riparius* growth (Growth)	Data not analysed because of complete adult mortality at positive control Station 5				
10 d *A. marina* cast formation (% lethality)	Kruskal Wallis (H)	0.34 (NS)	0.53 (NS)	14.4 (p<0.001)	No significant increases in lethality at Station 5 relative to Station 8
10 d *A. marina* cast formation (No. casts per animal)	Kruskal Wallis (H)	8.8 (p<0.01)	0.63 (NS)	16.1 (p<0.001)	Significant decreases in cast formation at Station 5 relative to Station 1
28 d *M.edulis in situ* (Clearance rate)	Kruskal Wallis (H)	7.4 (p<0.01)	4.9 (NS)	472 (p<0.001)	Significant decreases in clearance rate at Station 5, relative to Station 8
10 d *C. volutator in situ* (% lethality)	Kruskal Wallis (H)	7.5 (p<0.01)	0.31 (NS)	0.31 (NS)	Significant increases in lethality at Station 5 relative to Station 8

3.3.2 Confirmatory analysis of the sensitivity and discrimination of bioassays

In this section all the data for bioassays were used to confirm the results of the initial assessment described in Section 3.3.1 with the aim of determining which methods were sufficiently sensitive and discriminatory to be suitable for receiving water monitoring. This analysis involved identifying which stations at particular sites were identified by the bioassay responses as being different from the relevant reference location. The reference locations for the freshwater River Aire and Willow Brook sites were Station 1 (Bell Busk Road) and Station 8 (Apethorpe) respectively (see Table 1.6). For the Tees and Mersey Estuary sites the reference locations were Station 8 and Station 1 (Perch Rock) respectively (see Table 1.6). Tables 3.43-3.46 show the results of the statistical analyses (3-way ANOVA for all sampling deployment occasions and 2-way ANOVA for each sampling/deployment occasion followed by Dunnettt's test) for each bioassay with further information on the statistical procedure being given in Section 2.3.

Freshwater bioassays

The data for the ECL bioassay showed a large response range at both stations (River Aire: 2.1-99 ECL Units, Willow Brook: 2.3-44.6 ECL Units). At the River Aire there were increases in response from low levels at the upstream Station 1 to high levels at the downstream Stations 6-8 and the test consistently discriminated the reference location from all the other stations (see Table 3.43). On the Willow Brook, Stations 2,5 and 6 were consistently identified as showing different responses from the reference location (see Table 3.44). In previous studies the test has discriminated between receiving water stations at a number of sites (Environment Agency 1997b). However as described in Section 3.3.1 there is uncertainty as to the nature of substances (organic load or chemical contaminants) responsible for the bioassay response in environmental samples.

The Microtox-acute bioassay showed limited toxic responses at both sites (>95.6% of the control at the River Aire and >61% of the control at the Willow Brook). No stations on the River Aire were consistently discriminated as showing greater toxicity than the reference Station 1 (see Table 3.43) while on the Willow Brook only Station 7 was consistently identified as showing greater toxicity than reference Station 8 (see Table 3.44). For most samples taken at the two sites the Microtox - acute test showed stimulation of light production (that is a hormetic response), rather than an inhibitory toxic response. This finding is consistent with surveys of the Detroit River, Lake St Clair and the St Lawrence River in Canada which have shown that, generally, surface water samples taken outside effluent discharge mixing zones have no or limited toxicity (that is <20% light inhibition relative to the control) in the Microtox acute bioassay (Ribo *et al* 1985, Kaiser *et al* 1988 a,b).

The 7 d *L. minor* growth inhibition bioassay appeared to be measuring eutrophication at certain stations rather than toxicant-induced effects (see Table 3.5). No stations on the River Aire were identified as showing greater toxicity than the reference Station 1 on either test occasion (see Table 3.43) while on the Willow Brook only Station 2 (time 1) and Stations 3 and 4 (time 2) were discriminated from reference Station 8 (see Table 3.44).

The combination of juvenile production and lethality endpoints in the 10 d *D. magna* reproduction test was able to discriminate between stations at both sites where there was a 20% increase in lethality from that in samples from the reference location Station 1. At the River Aire the lethality endpoint separated the stations into two groups with low lethality (<10%) at Stations 1, (2), 3, 4 and 6 and higher levels of lethality (>32%) at Stations 5 (the positive control), 7 and 8. Lethality of *D. magna* was low at all stations on the Willow Brook but the juvenile production endpoint was able to discriminate between Station 1 and the other stations. The responses measured for the *D. magna* reproduction test on the River Aire were influenced by the flow conditions with a marked reduction in lethality being seen at downstream stations during sampling following a rain event. No lethal effects were evident in *D.magna* at either site after 2 days indicating that acute immobilisation tests on receiving water samples from these locations would not be sufficiently sensitive to identify impacts.

The 10 d *C. riparius* lethality bioassays discriminated between the reference location and other stations for both the River Aire and Willow Brook (see Tables 3.43 and 3.44). On the first sampling occasion on the River Aire (time 2) Stations 5 and 8 were discriminated from the reference location. However, discrimination on the second sampling occasion (time 3) was compromised by the unexpectedly high lethality at the reference location, which may have been due to a localised pollution episode. On the Willow Brook Station 1 was consistently discriminated from reference location Station 8.

The lethality endpoint of the 28 d *C. riparius* growth/lethality bioassay proved to be more sensitive and discriminatory than the 10 d *C. riparius* lethality bioassay on the first sampling occasion (time 2) (see Table 3.43). The endpoint discriminated between the reference location and Stations 2,5 and 6 which was not discriminated based on short-term lethality (see Figures 3.10 and 3.12). Of the other sublethal endpoints used, *Chironomus* development was the only one which discriminated between stations in a biologically meaningful way and added power to the discriminatory capacity of the chronic lethal endpoint. The growth endpoint for the 28d *C. riparius* bioassay gave ambiguous results and the biological significance of growth, as measured in these test systems is questionable, given that growth increased, rather than decreased, with increasing lethality. This finding is probably the result of increased food availability in test vessels where organisms have died. The test failed on the second occasion (time 3) with large variability in responses being observed. Bioassays used for receiving water monitoring would need to be reliable and further work is necessary to improve the performance of the 28d *C. riparius* growth/lethality bioassay.

For the freshwater systems, the 6 d *G. pulex* feeding rate *in situ* bioassay was capable of discriminating between stations on both occasions using a combination of the lethality and feeding rate endpoints (see Tables 3.43 and 3.44). On the River Aire the lethality endpoint discriminated between the reference location Station 1 and Stations 4,5 and 8 on both occasions.

The lethality-based 6 d *Daphnia magna in situ* and 10 d *Chironomus riparius in situ* bioassays were both capable of discriminating between stations but only on the first of the two deployment occasions (see Table 3.43). The discrimination of the *Chironomus* and *Daphnia in situ* bioassays should, however, be improved by increasing the number of test vessels used per station.

Table 3.43 Summary of the results of statistical analyses carried out on the bioassays used on the River Aire

Site	Test	ANOVA variance ratio (and significance) for different sampling/deployment times					Stations showing significant toxic response compared to those at reference control at different times				
		All	Time 1	Time 2	Time 3	Time 4	All	Time 1	Time 2	Time 3	Time 4
River Aire	ECL	276 (p<0.001)	660 (p<0.001)	306 (p<0.001)	321 (p<0.001)	218 (p<0.001)	2-8	3-8	2-8	2-8	2-8
	Microtox	5.75 (p<0.001)	94.8 (p<0.001)	37.2 (p<0.001)	6.44 (p<0.001)	125 (p<0.001)	6	6,7,8	None	3,5	6
	L. minor growth	6.35 (p<0.001)	1.24 (NS)	2.38 (p<0.01)	-	-	None	None	None	-	-
	D.magna reproduction - lethality	97.1 (p<0.001)	55.8 (p<0.001)	3.9 (p<0.01)	58.1 (p<0.001)	139 (p<0.001)	2,5,7,8	5,7,8	2	5,7,8	5,7,8
	D.magna reproduction - juvenile production	39.1 (p<0.001)	1.8 (NS)	1.3 (NS)	5.1 (p<0.001)	2.1 (NS)	None	None	None	None	None
	C. riparius lethality	22.3 (p<0.001)	-	4.16 (p<.0.01)	63.9 (p<0.001)	-	5,8	-	None	5,8	-
	C. riparius growth - lethality	8.97 (p<0.001)	-	8.97 (p<0.001)	Test failed	-	2,5,8	-	2,5,8	Test failed	-
	C. riparius growth - development	13.3 (p<0.001)	-	13.3 (p<0.001)	Test failed	-	4,8	-	4,8	Test failed	-
	C. riparius growth - growth	2.31 (NS)		2.31 (NS)	Test failed	-	-	-	-	Test failed	-
	G. pulex in situ - lethality	8.79 (p<0.01)	-	4.31 (p<0.01)	4.98 (p<0.01)	-	4,5,6,8	-	4,5,6,8	4,5,8	-
	G. pulex in situ - feeding rate	11.46 (p<0.001)	-	24.0 (p<0.001)	67.9 (p<0.001)	-	2,4-6,8	-	8	2,4-6,8	-
	D.magna in situ lethality	11.1 (p<0.001)	-	7.2 (p<0.05)	1.4 (NS)	-	3,5,7,8	-	3,5,7,8	None	-
	C. riparius in situ lethality	6.2 (P<0.001)	-	20.3 (p<0.001)	2.7 (NS)	-	3-5,7,8	-	5	None	-

Key: Times 1, 2, 3, 4 = 10/7/96, 9/8/96, 9/9/96, 17/10/96

Table 3.44 Summary of the results of statistical analyses carried out on the bioassays used on the Willow Brook

Site	Test	ANOVA variance ratio (and significance) for different sampling/deployment times					Station showing significant toxic response compared to those at reference control at different times				
		All	Time 1	Time 2	Time 3	Time 4	All	Time 1	Time 2	Time 3	Time 4
Willow Brook	ECL	75.9 (p<0.001)	82.7 (p<0.001)	830 (p<0.001)	59.7 (p<0.001)	-	2,5,6	2,5,6	2,5,6	2-8	-
	Microtox	90.5 (p<0.001)	321 (p<0.001)	285 (p<0.001)	325 (p<0.001)	-	7	7	7	7	-
	L. minor growth	9.95 (p<0.001)	4.04 (p<0.001)	3.25 (p<0.05)	-	-	2,3,4,7	2	3,4	-	-
	D.magna reproduction - lethality	0.78 (NS)	-	-	-	-	None	-	-	-	-
	D.magna reproduction - juvenile production	7.62 (p<0.001)	35.9 (p<0.001)	4.1 (p<0.001)	0.91 (NS)	-	1	1,2	None	None	-
	C. riparius lethality	8.82 (p<0.001)	-	21.1 (p<0.001)	7.95 (p<0.001)	-	1	-	1,2,3,5	1	-
	G. pulex in situ - lethality	2.84 (NS)	-	-	-	-	None	-	-	-	-
	G. pulex in situ - feeding rate	20.1 (p<0.001)	-	6.11 (p<0.001)	16.9 (p<0.001)	-	7	-	7	1,2,4,5, 7	-

Key: Times 1, 2, 3, = 24/7/96, 22/8/96, 20/9/96

Estuarine bioassays

At the estuarine sites all the laboratory-based water column bioassays showed discrimination between the stations (see Tables 3.45 and 3.46). The rapid ECL bioassay showed a large response range at both sites (5.1-733 ECL Units for the Tees Estuary and 1.2-149 ECL Units for the Mersey Estuary) and, in most instances, discriminated the reference location from all other stations.

The Microtox-acute bioassay also showed a large toxic response range at both sites (>11.2% of control for the Tees Estuary and >48.5% of the control for the Mersey Estuary), though responses at some stations indicated stimulation of light production (that is a hormetic response). The bioassay consistently discriminated Station 5 from the reference location on the Tees Estuary and Station 8 from the reference location on the Mersey Estuary (see Tables 3.45 and 3.46). Other stations at each site were discriminated at different sampling times depending on the tidal state.

The 24 h oyster embryo-larval development bioassay showed discrimination between stations (see Tables 3.45 and 3.46) and a large response range at both sites (PNR values of -12.6 to 100 for the Tees Estuary and -2.4-94.1 for the Mersey Estuary). However, at both sites the OEL bioassay showed an all or nothing response with either low PNR (<20%) or high PNR (>73%) values being measured. The extent of the responses and the discrimination between stations was dependent on the sampling time and tidal state (see Figures 3.18-3.19).

The combination of juvenile production and lethality endpoints in the 9 d *T. battagliai* reproduction bioassay was able to discriminate between stations at both estuarine sites (see Tables 3.45 and 3.46). At the Tees Estuary the lethality endpoint separated the stations into two groups with low lethality at Stations 1, 2, 4, 7 and 8 and complete lethality at Station 5 (the positive control) on all occasions and Stations 3 and 6 depending on the tidal state (see Figure 3.20). On the Mersey Estuary lethality of *T. battagliai* was generally low at Stations 1, 2, 3, 4 and 5 (< 28%) but was always 100% at Station 8 (see Figure 3.22). At both sites the juvenile production endpoint was not able to routinely discriminate between stations showing low adult mortality. However, this was because the number of juveniles produced was lower (Tees Estuary: 4-52%, Mersey Estuary: 1-22%) than levels in the seawater controls (see Table 3.19) indicating sub-lethal effects at all stations not considered acutely toxic.

The 10 d *C. volutator* lethality bioassay discriminated between the reference location and other stations for both the Tees and Mersey Estuary sites (see Table 3.45 and 3.46). On the Tees Estuary the bioassay discriminated between the reference Station 8 and Stations 1-6 and 2,3,5 and 6 on the first (time 2) and second (time 3) test occasions respectively. The results obtained for Tees Estuary samples on the two deployment occasions were consistent with those measured in previous surveys carried out by the Environment Agency in 1994 and 1995 (see Table 3.21). On the Mersey Estuary the bioassay discriminated between reference Station 1 and Stations 6-8 and 2 on the first (time 2) and second (time 3) test occasions respectively.

The lethality responses of the 28 d *C. volutator* bioassays on Tees Estuary samples were similar to those measured in the 10 d *C. volutator* bioassay (see Table 3.45). The discriminatory capability of the two bioassays based on lethality was also similar with the 28 d bioassay separating Stations 1,2,4, 5 and 6 from the reference Station 8, whereas the 10 d

bioassay discriminated Stations 1-6 from Station 8. For the Mersey Estuary samples the lethality responses in the 28 d bioassay were greater than those measured in the 10 d bioassay, but the 28 d bioassays was less discriminatory and no stations were identified as showing greater toxicity than the reference Station 1.

The 10 d cast formation response in *A. marina* on the Tees Estuary had a similar discriminatory capability as the 10 d *C. volutator* lethality bioassay (separating Stations 2,3,5, and 6 from reference Station 8), though the lethality response in the *A. marina* bioassay was less sensitive and discriminatory (see Table 3.45).

The 28 d *M. edulis* clearance rate *in situ* test discriminated between the reference locations on the Tees and Mersey estuaries and certain other stations on both deployment occasions (see Tables 3.45 and 3.46). On the Tees Estuary all the stations except Station 6 were identified as showing greater toxicity than the reference location on the first occasion while only Station 2 was discriminated on the second deployment occasion. On the Mersey Estuary the only station discriminated from the reference Station 1 on both deployment occasions was Station 8.

The 10 d *C. volutator* lethality *in situ* bioassay discriminated between the reference Station 8 and Stations 1-3 and 5 on both deployment occasions on the Tees Estuary (see Table 3.46). However, the discrimination of the bioassay on the second deployment occasion was reduced because the cages deployed at Stations 4 and 7 were lost due to tidal action. Improvements in the deployment procedure for this bioassay need to be made to minimise the loss of test vessels.

3.3.3 Summary of data on bioassay characteristics

On the basis of the data generated on sensitivity and discrimination of the bioassays used at the freshwater and estuarine sites a number of methods have been identified which are considered to be potentially suitable for receiving water monitoring. These methods are highlighted in Tables 3.47 and 3.48 and represent bioassays which, on the basis of the data generated in this study, are capable of distinguishing between unimpacted reference locations and stations subject to chemical contamination.

In this study tests with combined lethal and sub-lethal endpoints were shown to be advantageous because they can discriminate between stations at a site on two levels. In tests such as the laboratory-based water column *D. magna* and *T. battagliai* reproduction tests and the *Gammarus pulex in situ* test stations exposed to high concentrations of contaminants showed lethal effects while stations exposed to lower chemical concentrations showed effects on sub-lethal endpoints such as juvenile production and feeding rate.

At all the locations most of the bioassays which discriminated between stations also showed temporal variability in the results obtained. However, this is not unexpected given the dynamic nature of the environments being monitored. Changes in response over time, which also occur for macroinvertebrate survey data, reflect changes in:

- contaminant release from point sources and diffuse inputs;
- the available dilution in a receiving water system due to changes in river volume or tidal state.

The issue of temporal variability needs to be considered when determining the type of bioassays which will be deployed in a particular operational role (for example general quality monitoring or local environmental impact assessment) and this issue is considered further in Section 5. The issue of small scale spatial variability at a particular location also became evident in the analysis of the results since the effects are more pronounced for certain tests. Section 5 also considers how the issue of small scale spatial variability could be addressed in receiving water monitoring.

Table 3.45 Summary of the results of statistical analyses carried out on the bioassays used on the Tees Estuary

Site	Test	ANOVA variance ratio (and significance) for different sampling/deployment times					Station showing significant toxic response to that at reference control at different times				
		All	Time 1	Time 2	Time 3	Time 4	All	Time 1	Time 2	Time 3	Time 4
Tees Estuary	ECL	146 (p<0.001)	862 (p<0.001)	421 (p<0.001)	514 (p<0.001)	3731 (p<0.001)	1-7	2-5	1-7	1-7	1-7
	Microtox	1284 (p<0.001)	2437 (p<0.001)	1705 (p<0.001)	8336 (p<0.001)	1129 (p<0.001)	5,6	3,5	3,5	5,6	5,7
	OEL	1041 (p<0.001)	1613 (p<0.001)	1705 (p<0.01)	1284 (p<0.001)	1129 (p<0.001)	3-7	3-5	2-7	5-7	5-7
	T. battagliai reproduction - lethality	83.7 (p<0.001)	70.2 (p<0.001)	125 (p<0.001)	81.3 (p<0.001)	3.35 (p<0.01)	3-6	3-5	5	1,5,6	4,5
	T. battagliai reproduction - juvenile production	5.8 (p<0.001)	2.1 (NS)	3.4 (p<0.01)	0.65 (NS)	9.3 (p<0.001)	3,4,6	None	3	None	None
	C. volutator lethality	71.1 (p<0.001)	-	84.2 (p<0.001)	34.3 (p<0.001)	-	1-6	-	1-6	2,3,5,6	-
	C. volutator growth - lethality	217 (p<0.001)	-	217 (p<0.001)	Test failed	-	1,2,4-6	-	1,2,4-6	Test failed	-
	C. volutator growth - growth	29.7 (p<0.001)	-	29.7 (p<0.001)	Test failed	-	2,4,7	-	2,4,7	Test failed	
	A. marina cast formation - lethality	3.84 (p<0.01)	-	0.98 (NS)	2.93 (p<0.05)	-	3,6	-	None	6	-
	A. marina cast formation - cast formation	23.3 (p<0.001)	-	10.9 (p<0.001)	12.3 (p<0.001)	-	2,3,5,6	-	3-6	2,5,6	-
	M.edulis in situ clearance rate	16.1 (p<0.001)	-	11.1 (p<0.001)	6.7 (p<0.001)	-	2	-	1-5,7	2	-
	C. volutator in situ lethality	327 (p<0.001)	-	71.7 (p<0.001)	11.3 (p<0.001)		1-3,5	-	1-5,7	1-3,5	-

Key: Times 1, 2, 3, 4 = 2/8/96, 30/8/96, 27/9/96, 24/10/96

Table 3.46 Summary of the results of statistical analyses carried out on the bioassays used on the Mersey Estuary

Site	Test	ANOVA variance ratio (and significance) for different sampling/deployment times					Stations showing significant toxic response to those at reference control at different times				
		All	Time 1	Time 2	Time 3	Time 4	All	Time 1	Time 2	Time 3	Time 4
Mersey Estuary	ECL	883 (p<0.001)	1154 (p<0.001)	3816 (p<0.001)	4236 (p<0.001)	-	2-8	3-8	2-8	2-8	-
	Microtox	36.5 (p<0.001)	116 (p<0.001)	618 (p<0.001)	380 (p<0.001)	-	6-8	4-8	2-8	8	-
	OEL	263 (p<0.001)	232 (p<0.001)	186 (p<0.01)	355 (p<0.001)	-	4,6,7,8	8	5,6,8	7,8	-
	T. battagliai reproduction - lethality	52.4 (p<0.001)	20.3 (p<0.001)	31.2 (p<0.001)	36.5 (p<0.001)	-	6-8	3,8	6,7,8	3,8	-
	T. battagliai reproduction - juvenile production	2.0 (NS)	-	-	-	-	None	-	-	-	-
	C. volutator lethality	5.13 (p<0.001)	-	6.62 (p<0.001)	3.68 (p<0.01)	-	2-5,7,8	-	6-8	2	-
	C. volutator growth - lethality	7.21 (p<0.001)	-	-	7.21 (p<0.001)	-	None	-	-	None	-
	C. volutator growth - growth	3.38 (p<0.01)	-	-	3.38 (p<0.01)	-	2	-	-	2	-
	M.edulis in situ - clearace rate	9.6 (p<0.001)	-	9.2 (p<0.001)	3.2 (p<0.01)	-	8	-	8	8	-

Key: Times 1, 2, 3 = 31/7/96, 28/8/96, 25/9/96

Table 3.47 Summary of the suitability of freshwater bioassays for receiving water monitoring based on sensitivity and discrimination

Test type	Bioassay	Comments on sensitivity and discrimination	Suitability of method for receiving water monitoring
Laboratory-water column	ECL	Sensitive and discriminatory between reference locations and contaminated stations, but uncertainty over cause of test response	Potentially suitable if the cause of the response can be clarified
	Microtox - acute	Relatively insensitive with most samples showing stimulatory (hormetic) rather than inhibitory responses. Limited discrimination between reference locations and contaminated stations	Not suitable
	7 d *L. minor* growth	Relatively insensitive with most samples showing stimulatory (hormetic) rather than inhibitory responses. Limited discrimination between reference locations and contaminated stations	Potentially suitable for measuring eutrophication effects
	10 d *D.magna* reproduction	Combined lethal and juvenile production endpoints are sensitive and discriminatory between reference locations and contaminated stations	Suitable
Laboratory - sediment	10 d *C. riparius* lethality	Sensitive and discriminatory between reference locations and contaminated stations	Suitable
	28 d *C. riparius* growth/ lethality	Lethality and development endpoints are sensitive and discriminatory between reference locations and contaminated stations. However, growth endpoint is questionable. The assessment of the bioassays was compromised as one of the two tests failed	Further work is needed to improve the reliability of the test
In situ	6 d *G. pulex in situ*	Combined lethal and feeding rate endpoints are sensitive and discriminatory between reference locations and contaminated stations	Suitable
	6 d *D.magna in situ*	Sensitive and discriminatory between reference locations and contaminated stations, but high temporal and/or local spatial variability. Results may be influenced by physico-chemical factors such as dissolved oxygen	Further work is needed to improve the reliability of the test
	10 d *C. riparius in situ*	Sensitive and discriminatory between reference locations and contaminated stations, but high temporal variability	Further work is needed to improve the reliability of the test

Table 3.48 Summary of the suitability of estuarine bioassays for receiving water monitoring based on sensitivity and discrimination

Test type	Bioassay	Comments on sensitivity and discrimination	Suitability of method for receiving water monitoring
Laboratory-water column	ECL	Sensitive and discriminatory between reference locations and contaminated stations, but uncertainty over cause of test response	Potentially suitable if the cause of the response can be clarified
	Microtox - acute	Sensitive but limited discrimination between reference locations and contaminated stations	Not suitable
	24 h Oyster embryo-larval development	Sensitive and discriminatory between reference locations and contaminated stations, but generally all or nothing response	Suitable
	9 d *T. battagliai* reproduction	Combined lethal and juvenile production endpoints are sensitive and discriminatory between reference locations and contaminated stations	Suitable
Laboratory - sediment	10 d *C. volutator* lethality	Sensitive and discriminatory between reference locations and contaminated stations	Suitable
	28 d *C. volutators* growth/ lethality	Lethality endpoint is sensitive and discriminatory between reference locations and contaminated stations. However, growth endpoint is questionable. The assessment of the bioassays was compromised as one of the two tests failed	Further work is needed to improve the reliability of the test
	10 d *A. marina* cast formation/lethality	Cast formation endpoint is sensitive and discriminatory between reference locations and contaminated stations. However, lethality endpoint is relatively insensitive and non-discriminatory	Suitable
In situ	28 d *M.edulis in situ*	Clearance rate endpoint is sensitive but relatively non-discriminatory between reference locations and contaminated stations	Not suitable
	10 d *C. volutator in situ*	Sensitive and discriminatory between reference locations and contaminated stations, but high temporal variability	Further work is needed to improve the reliability of the test

3.4 Power of the bioassay procedures

3.4.1 Introduction

Section 3.3 assessed the suitability of bioassays for receiving water monitoring based on their sensitivity and discrimination. However, it is also important that the bioassays when used in an appropriate test design are sufficiently powerful to be able to detect pre-defined differences in response between stations which are likely to equate to meaningful population changes (see Section 5).

The power of a bioassay to be able to detect differences in response between stations is determined by a range of factors including:

- the number of samples taken at a station or cages deployed at a station;
- the number of replicates of a sample tested in laboratory-based water column or sediment bioassays;
- the number of organisms used in each replicate of laboratory-based bioassays or in *in situ* bioassays;
- the inherent variability of the procedure.

Tables 3.49 and 3.50 summarise information on the power of the bioassays used at each site on the basis of minimum detectable differences (MDD). These were calculated using the information obtained in the 2-way analysis of variance carried out for the bioassay data obtained on each sampling/deployment occasion described in Section 3.1. An explanation of the procedures adopted to determine the minimum detectable differences is given in Section 2.3. In the calculation of the MDD the probability of a Type I statistical error α (that is the false conclusion that the responses measured in bioassays at the stations are different from those at the reference station) was set at 5% while the probability of making a Type II statistical error β (that is the false conclusion that there are no differences among the mean responses of bioassays at the reference and other stations) was set at 10%. These values are typically used when assessing the power of tests (Zar 1984).

3.4.2 Laboratory-based water column bioassays

Generally the laboratory-based water column bioassays (that is ECL, Microtox-acute, *L.minor* growth inhibition, OEL and the lethality endpoints of the *D. magna* and *T. battagliai* reproduction tests) were sufficiently powerful to be able to detect differences in response between stations of 20-25% or lower. In the case of the ECL and OEL bioassays differences in response of 5% between stations could be detected on all occasions. In contrast, the power of the *L. minor* growth inhibition test and the juvenile production endpoints of the *D. magna* and *T. battagliai* reproduction tests, as used in this study were low (that is at least 50% differences in response between stations were needed for statistical significance). The power of the juvenile production endpoint in the reproduction tests could be enhanced by increasing the number of organisms per sample. If, for example, 10 animals per sample replicate had been

used in the *D. magna* and *T. battagliai* reproduction tests instead of five it would be expected that (on the basis of previous unpublished observations) that the within group variability (within cells MS - see Section 2.3) would decrease and the test would be sufficiently powerful to then detect differences in response between stations of 35% or lower.

In the *D. magna* reproduction test the power of the test could also be enhanced by extending the duration of the bioassay from 10 to 14 days because of the increased number of offspring that would be produced in the bioassay under control conditions (Environment Agency 1999). Increasing the numbers of organisms per sample used would also improve the power of the lethality endpoint in the reproduction tests.

3.4.3 Laboratory-based sediment bioassays

In the laboratory-based sediment bioassays the lethality endpoints in the 10 d and 28 d *C. riparius* and *C. volutator* sediment bioassays were generally sufficiently powerful to detect differences of 20% or lower between stations. The lethality and cast formation of the *A. marina* were less powerful but the power of the cast formation could be improved by using additional organisms.

3.4.4 *In situ* bioassays

There was considerable variability in the power of the *in situ* bioassays which, to a large extent, reflected the degree of development of the methods. At the freshwater sites, the standardised *G. pulex* feeding rate/lethality bioassay had the greatest power and the feeding rate endpoint was able to detect differences of 23-39%, relative to values at the reference location. This level of power is consistent with previous studies using the same study design which were able to detect 25-30% differences in response between stations (Crane *et al* 1993).

The power of the *in situ D. magna* and *C. riparius* lethality bioassays was generally low (that is at least 50% differences in response between stations were needed for statistical significance), but this could be improved by increasing the number of cages deployed at a station (for example from 4 to 8)(see Section 3.4.2).

At the estuarine sites the *in situ C. volutator* lethality bioassay was sufficiently powerful to detect differences of 13% between stations when all the cages were retrieved but showed reduced power when cages were lost. The *M. edulis* clearance rate bioassay generally showed low power and was only able to detect greater than 47% differences in response, relative to those at the reference station. The low power reflected the variability in response of the deployed organisms at a station.

3.4.5 Approach for designing general quality assessment or local environmental impact assessment schemes

There are now software based methods for investigating the test method design required to achieve a pre-defined level of power (for example Query Advisor) and these can be used with a consideration of the available resources to achieve the most cost-effective sampling/testing strategy. This design may involve maximising the number of tests carried out at stations where

intermediate effects are expected compared to locations where no or complete effects are anticipated.

It is also more appropriate for the test design adopted for local environmental impact assessments to be more rigorous than for general quality assessment with greater number of replicates and test organisms being used in the laboratory-based bioassays or *in situ* deployments.

It is also important to attempt to link the minimum detectable difference that a bioassay should be able to detect with consequences for the population dynamics of the test species. In this way statistical differences in bioassay responses can be considered to have a greater degree of ecological relevance (see Section 5).

Table 3.49 Summary of the Minimum Detectable Differences for freshwater bioassays and the implications for the power of the test

Test, endpoint and location	Minimum detectable difference on each sampling/deployment occasion				Comments on power of bioassay
	Time 1	Time 2	Time 3	Time 4	
ECL - ECL Units: River Aire	1.0	1.0	1.1	1.2	In all instances the test design was sufficiently powerful to detect differences of response of 5% between stations
ECL - ECL Units: Willow Brook	1.0	1.0	1.2	-	
Microtox - % of control:River Aire	8.1	7.0	17.5	4.6	In most instances the test design was sufficiently powerful to detect differences of response of 10% between stations
Microtox - % of control: Willow Brook	7.2	8.3	6.2	-	
L.minor growth - growth rate					
River Aire	0.025	0.138	-	-	In most instances the test design was sufficiently powerful to detect differences of response of 25% or lower between stations
Willow Brook	0.08	0.077	-	-	
D.magna reproduction - % lethality					
River Aire	18.9	12.4	18.2	8.8	In all instances the test design was sufficiently powerful to detect differences of response of 20% or lower between stations
D.magna reproduction - No of young					
River Aire	7.3	8.2	7.5	6.5	Low power which could be improved by modification of test design (for example increasing the number of organisms per replicate)
Willow Brook	4.9	7.9	5.8	-	
C.riparius lethality - % lethality (10d)					
River Aire	-	12.6	8.8	-	In all instances the test design was sufficiently powerful to detect differences of response of 13% or lower between stations
Willow Brook	-	8.7	7.1	-	
C.riparius lethality - % lethality (28d)					
River Aire	-	3.9	Test failed	-	In the instances when the test was successful the test design was sufficiently powerful to detect differences of response of 10% or lower
G.pulex in situ - % lethality					
River Aire	-	28.0	42.5	-	Test design was sufficiently powerful to detect differences of response of 43% or lower between stations
Willow Brook	-	-	-	-	
G.pulex in situ - feeding rate					
River Aire	-	0.096	0.097	-	Test design was generally sufficiently powerful to detect differences of response of 23-39%, relative to those at the reference station
Willow Brook	-	0.113	0.07	-	
D.magna in situ - % lethality: River Aire	-	77.8	61.0	-	Low power which could be improved by modification of test designs (for example increasing the number of cages per station)
C.riparius in situ - % lethality: River Aire	-	37.5	88.2	-	

Table 3.50 Summary of the Minimum Detectable Differences for estuarine bioassays and the implications for the power of the test

Test, endpoint and location	Minimum detectable difference on each sampling/deployment occasion				Comments on power of bioassay
	Time 1	Time 2	Time 3	Time 4	
ECL - ECL Units: Tees Estuary	3.1	1.0	1.0	1.0	In all instances the test design was sufficiently powerful to detect differences of response of 5% between stations
ECL - ECL Units: Mersey Estuary	1.0	1.0	1.0	-	
Microtox - % of control: Tees Estuary	4.2	2.8	2.0	3.4	In all instances the test design was sufficiently powerful to detect differences of response of 10% between stations
Microtox - % of control: Mersey Estuary	8.1	3.5	5.1	-	
OEL - PNR: Tees Estuary	1.1	1.4	1.0	1.3	In all instances the test design was sufficiently powerful to detect differences of response of 5% between stations
OEL - PNR: Mersey Estuary	1.6	3.2	1.6	-	
T.battagliai reproduction - % lethality					
Tees Estuary	14.9	5.5	13.2	61.0	In most instances the test design was sufficiently powerful to detect differences of response of 25% or lower between stations
Mersey Estuary	27.0	23.0	15.7	-	
T.battagliai reproduction - No of young					
Tees Estuary	5.9	3.9	8.7	7.5	Low power which could be improved by modification of test design (for example increasing the number of organisms per replicate)
C.volutator lethality - % lethality (10d)					
Tees Estuary	-	12.5	21.1	-	In all instances the test design was sufficiently powerful to detect differences of response of 22% or lower between stations
Mersey Estuary	-	6.9	10.5	-	
C.volutator lethality - % lethality (28d)					
Tees Estuary	-	2.4	Test failed	-	In all instances the test design was sufficiently powerful to detect differences of response of 23% or lower between stations
Mersey Estuary	-	-	22.5	-	
A.marina - % lethality: Tees Estuary	-	44.1	69.1	-	Lethality endpoint has low power, power of cast formation endpoint could be improved by modified of test design
A.marina - % cast formation: Tees Estuary	-	13.0	12.9	-	
M.edulis in situ - clearance rate					
Tees Estuary	-	1.04	1.21	-	Low power due to high inherent response variability
Mersey Estuary	-	1.41	1.66	-	
C.volutator in situ - % lethality: Tees Estuary	-	12.5	29.9	-	Sufficiently powerful to detect differences of 20% or lower when no cages were lost

3.5 Comparison of ecotoxicological data and biological survey information

3.5.1 Summary of the biological data obtained at different stations

River Aire

The commonly used univariate biological measures; Biological Monitoring Working Party (BMWP) score, Average Score Per Taxon (ASPT) and number of taxa for the stations sampled on the River Aire are summarised in Table 3.51. The ranking of stations on the basis of each of these univariate measures (highest to lowest) was the same, namely 1>2>5>4>8. The reference Station 1 showing the highest BMWP (149) and ASPT (5.52) values and number of taxa (27) before the river showed a marked decline in biological quality by Station 4 (BMWP= 27, ASPT = 3.0, number of taxa = 9) which was maintained at Stations 5 (positive control) and 8 (see Table 3.51).

Table 3.51 Summary of biological survey data for stations on the River Aire (ND-No data)

Biological index	Values for different biological indices at different stations							
	1	2	3	4	5	6	7	8
BMWP	149	68	ND	27	47	ND	ND	22
ASPT	5.52	3.78	ND	3.0	3.36	ND	ND	2.75
No. of taxa	27	18	ND	9	14	ND	ND	8

Table 3.52 summarises the abundance data for each station on the River Aire used in the comparison. Station 1 showed a high number of pollution sensitive organisms (such as the mayfly family Ephemeridae) with pollution insensitive species such as the isopod crustacean *Asellus aquaticus* and oligochaetes worms becoming more dominant at Stations 5 and 8, indicating impacted locations.

The biological data from sites 1, 2, 4, 5 and 8 were analysed using non-metric multiple dimensional scaling (MDS) based on Bray-Curtis similarities calculated using 4th root transformed abundance data. This transformation was applied for biological reasons, namely to downweight the influence of the high abundance values in the calculation of between sample similarities. The resulting 2-dimensional configuration for the macroinvertebrate abundance data is illustrated in Figure 3.47. It is evident that Station 1 and the next Station downstream (2) are contrasted with the other sites along the x-axis reflecting the pollution gradient revealed by the biotic indices. However, the multivariate analysis reveals that Station 5 is distinct from all of the other stations. It is most distant from Stations 1 and 2 but is also separated from Stations 4 and 8 along the y-axis indicating a difference in community composition. The reason for the change in community composition at Station 5 is unclear though it may be the result of the input from the large Esholt sewage treatment works upstream of this station.

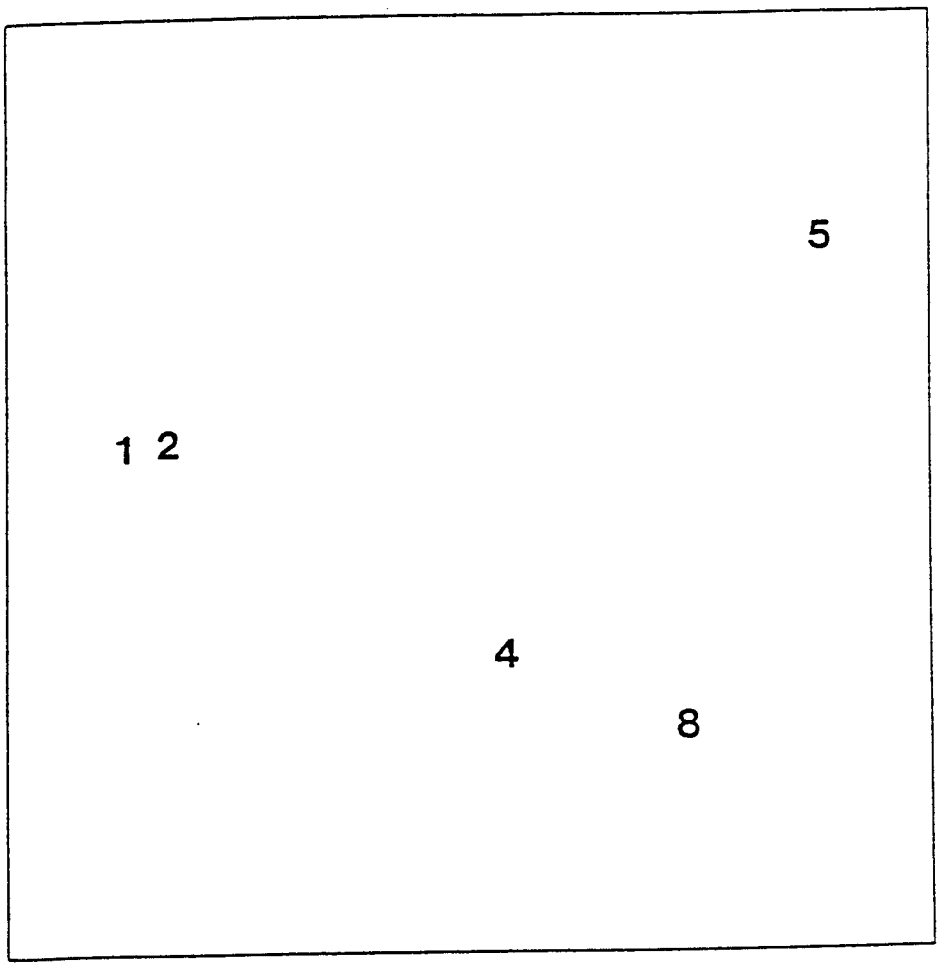

Figure 3.47 Two-dimensional MDS ordination of macroinvertebrate abundance data for stations on the River Aire (numbers indicate stations)

Table 3.52 Abundance data on organisms recorded at each station on the River Aire

Organisms present	Number of organisms at each station				
	1	2	4	5	8
Ancylus fluviatilis	35	345	0	38	0
Asellus aquaticus	1	2	675	10700	2405
Asellus meridianus	0	0	45	0	0
Athripsodes	10	0	0	0	0
Baetidae	70	0	0	0	0
Baetis	0	5	0	0	0
Baetis rhodani	840	2	2	15	0
Bithynia leachi	0	0	0	0	0
Bithynia tentaculata	0	6	1	0	0
Caenis rivulorum	2	0	0	0	0
Ceratopoaonidae	1	0	0	0	0
Chironomidae	360	178	48	2	3
Dendrocoelidae	0	0	0	13	3
Dicranola	26	0	0	0	0
Dugesia	0	4	1	1	0
Dytiscid larva	1	0	0	0	0
Ecdynurus	5	0	0	0	0
Elmis aenea	130	1	0	0	0
Ephemera danica	5	0	0	0	0
Ephemerella ignita	61	0	0	0	0
Erpobdella	0	0	0	0	3
Erpobdella octoculata	6	25	68	1073	6
Gammarus pulex	9	0	0	0	0
Glossiphonia	0	0	0	0	1
Glossiphonia complanata	4	1	0	9	0
Glossiphonia heteroclita	0	0	0	0	1
Gyrinidae larva	1	0	0	0	0
Helobdella stagnalis	2	1	2	0	0
Hemerodramiinae	0	1	1	0	0
Hemiclepsis marginata	0	0	0	1	0
Hydracanna	5	3	0	0	0
Hydraena gracilis	6	0	0	0	0
Hydropsyche instabilis	14	0	0	0	0
Hydropsyche pellucidula	16	0	0	0	0
Hydropsyche siltalai	138	0	0	0	0
Leuctra fisca	13	0	0	0	0
Limnius volckmani	65	0	0	0	0
Lymnaea palustris	0	1	0	0	0
Lymnaea peregra	4	505	2	0	1
Oligochaeta	37	72	12	1195	1
Orectochilus villosus	1	0	0	0	0
Oulimnius	138	17	0	0	0
Physa fontinalis	0	6	0	4	0

Organisms present	Number of organisms at each station				
	1	2	4	5	8
Planorbis albus	5	11	0	110	0
Planorbis contortus	0	14	0	0	0
Planorbis crista	0	1	0	0	0
Planorbis vortex	0	5	0	3	0
Polycelis	0	0	0	18	0
Potamopyrgus jenkinsii	62	2180	0	8	0
Rhyacophila dorsalis	8	0	0	0	0
Riolus subviolaceus	1	0	0	0	0
Segmentina complanata	0	0	0	6	0
Sericostoma personatim	29	0	0	0	0
Simuliidae	690	195	0	0	0
Sphaeriidae	68	540	0	23	15
Tinodes waeneri	0	1	0	0	0
Tipola	0	4	0	0	0
Valvata cristata	0	7	0	0	0
Valvata macrostoma	0	17	0	0	0
Veliidae	1	0	0	0	0

Willow Brook

Table 3.53 shows the univariate biological measures of BMWP and ASPT measured at the sampling stations during previous surveys and the one conducted in this study.

The ranking of stations on the basis of either univariate measure (highest to lowest) using the data collected during this study (20/9/96) was 8>6>5>3>4>7>1>2 with reference Station 8 (BMWP = 136, ASPT = 4.86) and 6 (BMWP = 83, ASPT = 4.61) showing the highest values. The lowest values were measured at Stations 1 (BMWP = 19, ASPT = 3.17) and 2 (BMWP = 12, ASPT = 2.40) with intermediate values (BMWP = 28-60 and ASPT = 3.11-4.29) being recorded at other stations. The values obtained during this study were generally consistent with those measured earlier in the year (30/5/96) in that Stations 1 and 2 had the lowest values and Stations 6 and 8 the highest. The pattern for the others stations showed a degree of variability between the different sampling times.

Table 3.53 Summary of biological survey data for stations on the Willow Brook (ND-No data)

Biological index	Date	Values of for different biological indices at different stations							
		1	2	3	4	5	6	7	8
BMWP	1995	6	ND	ND	53	26	45	28	95
	30/5/96	9	12	30	50	54	61	41	138
	20/9/96	19	12	55	49	60	83	28	136
ASPT	30/5/96	3.00	2.40	3.75	4.17	4.15	4.36	3.24	4.93
	20/9/96	3.17	2.40	4.23	4.08	4.29	4.61	3.11	4.86

Tees Estuary

Six stations (1,2,4,5,7 and 8) were sampled in the Tees Estuary to assess biological quality and at each station five replicate cores were taken. Table 3.54 and Figures 3.48 and 3.49 show the summary statistics for the univariate biological measures (see Section 2.3.3) namely the primary and derived variables: total number of taxa (T), total number of individuals (A), total biomass (B), ratios of A:T and B:A and the number of nematodes and the diversity indices Shannon-Wiener (H'), Pielou's index of eveness (J') and Simpsons diversity index (1/D).

The plots of the primary variables (Figure 3.48) indicate highest and most variable total abundance (A) at Station 5 and lowest at Stations 7 and 8 (reference location). The apparent trend in total number of taxa (T) is the opposite of that observed for total abundance with fewest taxa at Station 5 and greater numbers at Stations 7 and 8. The pattern in total biomass (B) suggests consistent mean values at Stations 1, 2, 4 and 5 with reduced values at Stations 7 and 8. The derived measure A/T shows a distinct increase at Station 5 suggesting that the samples at this site are dominated in terms of abundance by relatively few taxa. Values of B/A are lowest at Stations 5 and 8 suggesting that the average size of the animals are smaller at these stations. The measures of diversity (H', J' and 1/D) (Figure 3.49) all suggest the same pattern with consistent values at Stations 1, 2, 4 and 7 a distinct reduction in diversity at Station 5 and greatest diversity at Station 8.

The response of the univariate measures suggest that Station 5 (positive control) is affected by pollution with macrobenthic communities of high and variable abundance and low diversity (that is relatively low numbers of taxa of smallish size dominated by relatively few species). The samples from Station 5 are dominated by *Capitella capitata*, an indicator of organic enrichment. Station 5 corresponds to the point of entry into the estuary of the Dabholm Gut culvert receiving effluent from the large ICI Wilton industrial complex.

Table 3.54 Summary of the biological survey data for stations on the Tees Estuary (ND-No data)

Biological index	Values for different biological indices at different stations							
	1	2	3	4	5	6	7	8
Total no. of taxa (T)	12	11.4	ND	15.4	7.4	ND	14.4	25.2
Total no. of individuals (A)	1398	931	ND	1175	2378	ND	552	562
Total biomass (B)	2283	1938	ND	1952	1965	ND	959	419
A/T	122	65	ND	76	318	ND	39	25
Number of nematodes	0.8	33	ND	138	0.2	ND	1	2
B/A	1.8	1.8	ND	2	1	ND	0.6	2
Shannon-Weaver diversity index (H')	1.46	1.34	ND	1.73	0.318	ND	1.41	2.18
Pelou's evenness index (J')	0.59	0.51	ND	0.63	0.155	ND	0.54	0.7
Shannon diversity index (1/D)	0.31	0.19	ND	0.23	0.88	ND	0.32	0.15

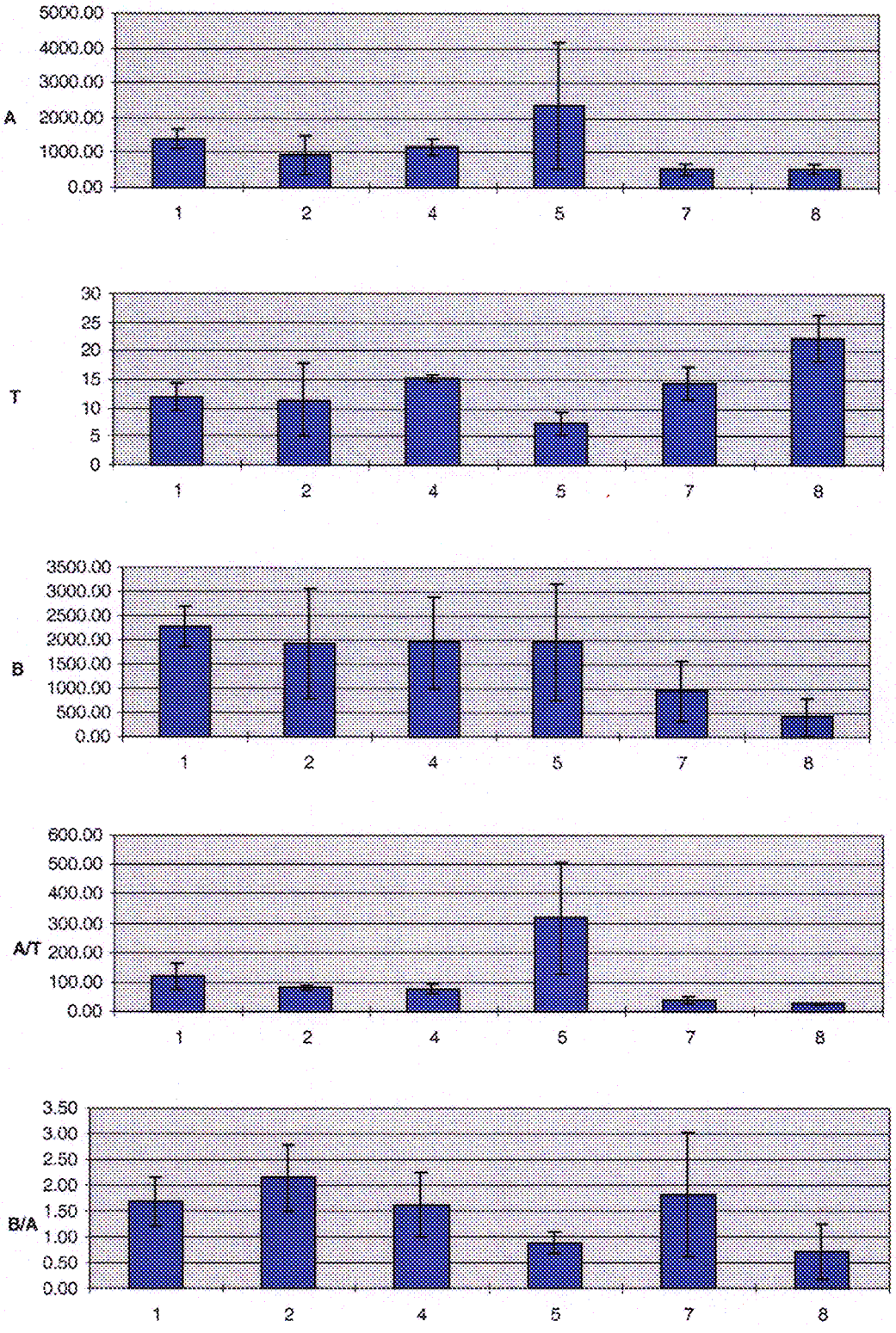

Figure 3.48 Primary and derived univariate biological measures of community structure at selected stations on the Tees Estuary (means ± SD, A = abundance, T = total number of taxa, B = biomass as mg, A/T and B/A = ratios of A:T and B:A respectively)

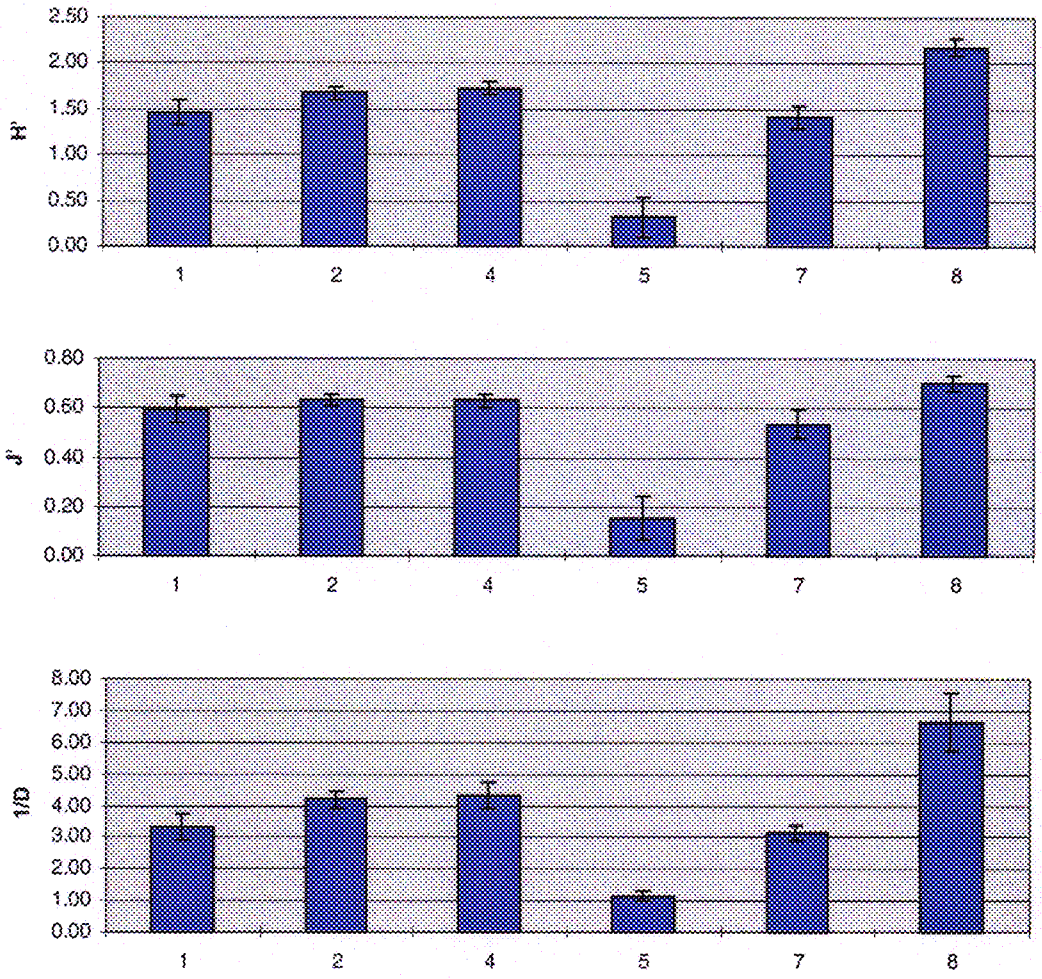

Figure 3.49 Univariate measures of biological diversity at selected stations on the Tees Estuary (means ± SD, H' = Shannon-Wiener diversity index, J' = Pielou's index of eveness, 1/D = Simpson's diversity index)

Figure 3.50 shows the MDS ordination patterns for the macro-invertebrate abundance and biomass data in Table 3.55. The patterns observed were similar for both parameters. Stations 1, 2 and 4 were separated from Stations 5, 7 and 8 on one dimension while on the other dimension Stations 5 and 8 were separated from Stations 1, 2, 4 and 7. The exact reasons for the changes in community composition at the different stations is unclear.

Table 3.55 Summary of the abundance and biomass data for organisms present at the stations on the Tees Estuary

Organisms	Abundance data at a station (mean number)						Biomass data at a station (mean mg wet weight)					
	1	2	4	5	7	8	1	2	4	5	7	8
Pholoe sp.	0	0	0	0	0	2.4	0	0	0	0	0	0
Eteone longa	1.8	1.8	6	0	3.6	0.6	17	13.0	47.2	0	14.5	0.98
Phyllodoce mucosa	0	0.8	3.6	0	1.2	0.2	0	8.1	39.2	0	14.7	1.34
Eumida bahusiensis	0	0.2	0.6	0	0	0	0	0.12	0.34	0	0	0
Microphthalmus sczelkowii	0	0	0.6	1.2	0.2	0	0	0	0.18	0.24	0.06	0
Nereis diversicolor	41.4	14.2	16.8	0	7	0.4	1618	600.9	1182	0	544.6	2.96
Nereis virens	0.4	4.2	0.4	0	0	0.4	16.4	515.2	21.3	0	0	173.9
Nephtys hombergii	0	0	0	0	0.6	1.4	0	0	0	0	103.5	64.5
Ophryotrocha hartmanni	0.2	5	3.2	0	4.8	65.6	0.02	0.5	0.32	0	0.48	6.58
Scoloplos armiger	0	0	0	0	0	0.4	0	0	0	0	0	0.2
Malacoceros fuliginosus	22.8	48.2	130.2	7.6	8.8	0.2	28.1	128.4	234.6	21.4	25.1	0.12
Polydora ciliata	224	199	230.2	84.4	0.8	3.4	209.5	186.3	103.7	42	0.94	1.66
Pygospio elegans	1.4	0.2	1.2	0	0.6	0.6	0.58	0.18	0.9	0	0.96	0.22
Spio martinensis	0	0	0	0	6.8	0.2	0	0	0	0	3.48	0.08
Streblospio shrubsolii	0	0	0	0.4	2.2	96.4	0	0	0	0.14	1.34	37.5
Chaetozone gibber	0	0	0	0	0	0.4	0	0	0	0	0	0.68
Cirriformia tentaculata	0.2	0.2	1.6	0	0	0	0.16	4.96	10.6	0	0	0
Tharyx vivipara	0	0	0	0	0	21	0	0	0	0	0	12.9
Capitella capitata	4.6	212.4	413.8	2158	256	8.6	6.6	302.2	189.1	1871	122.4	4.1
Mediomastus fragilis	0	1.2	0	0	0	18.2	0	1.54	0	0	0	8.34
Euchone sp. A	0	0	0.2	0	0	10	0	0	0.62	0	0	4.68
Fabricia sabella	0.8	2.2	3.6	0	0	0.2	0.16	0.42	0.7	0	0	0.04
Manayunkia aestuarina	650.2	140	250	0.2	5.6	54.8	63.7	13.7	24.5	0.02	0.56	5.3
Paranais sp. A	0	0	0	48.4	0	0.2	0	0	0	0	0	0.02
Paranais sp. B	0	0.4	0	0	0	1.2	0	0.04	0	4.84	0	0.12
Tubificoides swirencoides	0	0	0	0	0	117.8	0	0	0	0	0	35
Tubificoides cf. galaciensis	0	0	0	0	0	104.2	0	0	0	0	0	30.9

Organisms	Abundance data at a station (mean number)						Biomass data at a station (mean mg wet weight)					
	1	2	4	5	7	8	1	2	4	5	7	8
Tubificoides benedii	38.6	25.8	60	5.6	99	34.4	42.2	28.2	64.1	6.02	75.2	14.8
Tubificoides pseudogaster	116.2	258.8	28.6	13	148.6	4.2	52.0	115.9	17.5	5.78	37.5	1.06
Enchytraeidae indet.	0.2	0	0	54	0.4	1	0.04	0	0	10.8	0.22	0.24
Talitridae indet.	0	0	0	0	0.2	0	0	0	0	0	0.52	0
Corophium insidiosum	0	0.4	0.2	0	0	0	0	0.14	0.08	0	0	0
Hydrobia ulvae	292.2	14.8	22	4.8	2.8	1.8	146.8	7.44	9.5	2.64	1.32	1.1
Bivalve juv. indet.	0	0	0	0	0.2	0	0	0	0	0	0.06	0
Mytilus edulis	0.2	0	0.4	0	0.4	2.6	0.08	0	0.12	0	0.16	0.82
Mysella bidentata	0	0.8	0.6	0	0.8	0.4	0	4.46	4.18	0	4.68	0.8
Cerastoderma edule	1.4	0.4	1	0	0.6	1	2.14	0.88	2.3	0	6.76	0.42
Fabulina fabula	0	0	0	0	0.6	0.8	0	0	0	0	0.16	0.12
Macoma balthica	0.4	0	0	0	0	0.4	78.4	0	0	0	0	0.14
Abra alba	0	0	0	0	0	7	0	0	0	0	0	6.58
Mya juv. indet.	0.8	0.2	0	0	0	0	1.16	4.86	0	0	0	0

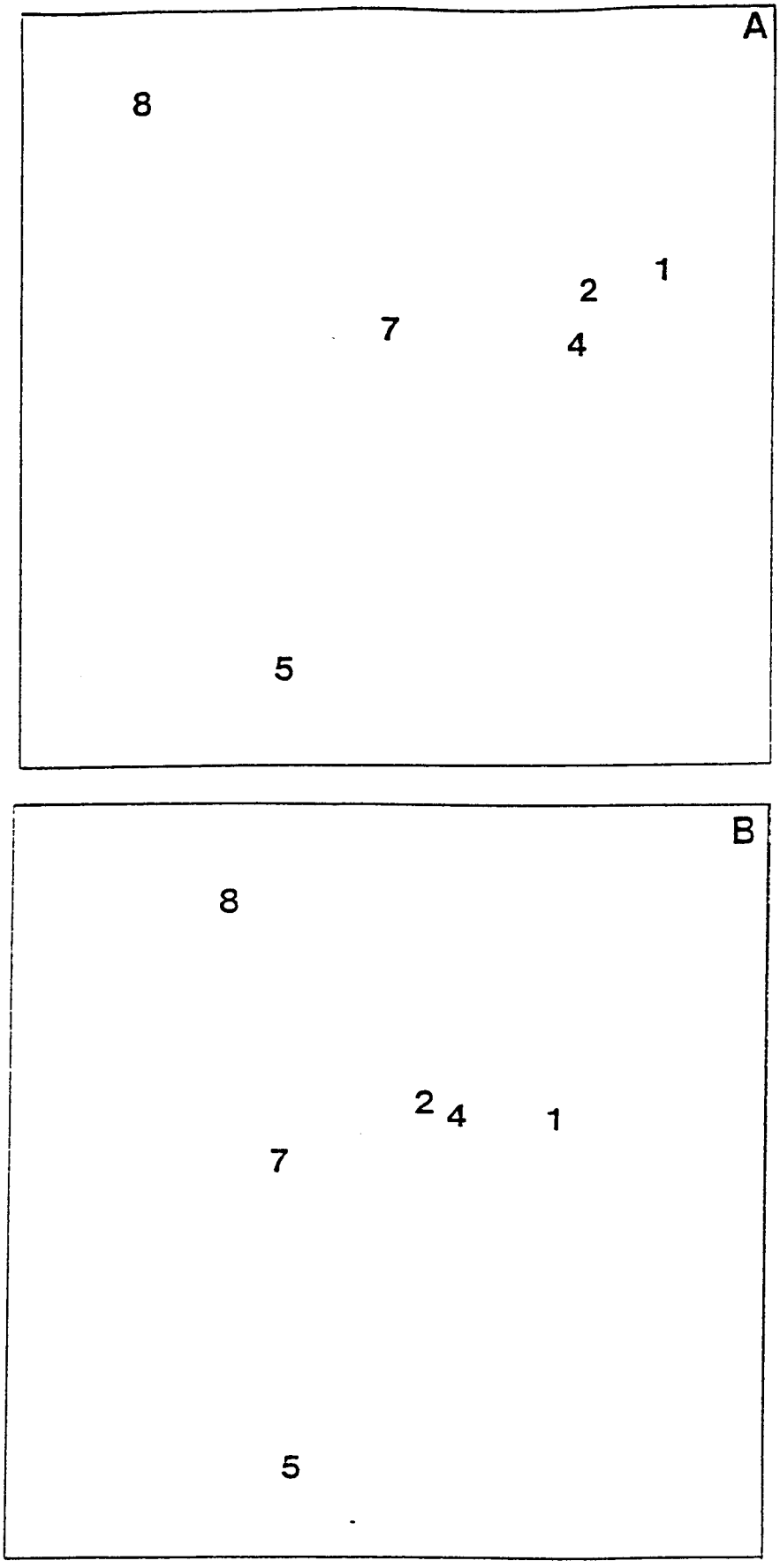

Figure 3.50 Two-dimensional MDS ordination of macroinvertebrate abundance (A) and biomass (B) data for stations on the Tees Estuary (number indicates station)

Mersey Estuary

Table 3.56 shows the biological survey data obtained for the Mersey Estuary. Information was only obtained at four stations due to logistical problems (see Section 2.2.7). On the basis of the number of taxa recorded the stations were ranked (clean to impacted) 3>4>6>8 while the Shannon-Weaver diversity index (H') ranked the stations 3>4>8>6.

Table 3.56 Summary of the biological survey data for stations on the Tees Estuary (ND-No data)

Biological index	Values for different biological indices at different stations							
	1	2	3	4	5	6	7	8
Total no. of taxa (T)	ND	ND	12.8	6.8	ND	5.8	ND	5.0
Total no. of individuals (A)	ND	ND	1948	292	ND	1036	ND	4123
Total biomass (B)	ND	ND	590	110	ND	580	ND	1210
A/T	ND	ND	152	43	ND	179	ND	844
B/A	ND	ND	0.31	0.38	ND	0.61	ND	0.29
Shannon-Weaver diversity index (H')	ND	ND	1.57	0.94	ND	0.62	ND	0.69
Pelou's evenness index (J')	ND	ND	0.62	0.48	ND	0.38	ND	0.44
Shannon diversity index (1/D)	ND	ND	0.26	0.51	ND	0.66	ND	0.57

3.5.2 Relationship of ecotoxicological data to biological survey data

In the riverine and estuarine studies carried out in this project one of the key objectives was to assess whether the results of bioassays provided data which was complementary to that from biological surveys or which resulted in the same conclusions being drawn as those from the biological surveys (that is indicating redundancy). In the case of the River Aire and Tees Estuary sites strong correlations between bioassay responses and macro-invertebrate survey data were expected because historical information indicated that chemical contaminants were responsible for the gradients in biological quality observed at the two sites (see Section 1.4).

At a site the extent of the agreement between the information from the two data sets will depend on a series of factors including whether:

- uniform conditions exist across the locations at a site so that any differences measured in benthic macroinvertebrate surveys are more likely to be due to the presence of chemical contaminants rather than habitat suitability;

- resident benthic macro-invertebrate populations at the different locations show differences in sensitivity due to factors such as organism size. Several studies have shown a strong inverse relationship between the sensitivity of

organisms to contaminants and body size (Gauss *et al* 1985, Kosalwat and Knight 1987);

- acclimation or adaptation of benthic macroinvertebrate populations or communities has occurred at given locations at a site. Because of the ability of natural populations to acclimate or adapt to contaminants the results of the bioassays may overestimate the effects observed in the field. However, in contrast, these populations acclimated or adapted to certain contaminants may be more vunerable to other stressors which may mitigate potential differences in response seen in the two monitoring approaches .

River Aire

Table 3.57 summarises the results from the Spearman's rank correlations between the macroinvertebrate abundance and bioassay data for the River Aire. The best single indicator of natural macroinvertebrate abundance in the River Aire was *in situ* lethality of the water column invertebrate *D. magna* (Spearman's rank correlation=0.752). Other predictors with high Spearman's rank correlations were laboratory tests with *C. riparius* (lethality) and the *G. pulex in situ* feeding rate bioassay. Tests with ECL, Microtox, *L.minor* growth inhibition, *D. magna* reproduction in the laboratory and *C. riparius* lethality *in situ* were poorer predictors of macroinvertebrate abundance patterns.

Table 3.57 Spearman's rank correlations between macroinvertebrate abundance and bioassay results for the River Aire

Explanatory variable	Spearman's rank correlation
Laboratory-based water column bioassays	
Enhanced chemiluminescence (ECL)	0.306
Bacterial bioluminescence (Microtox)	-0.413
7 d *L.minor* growth inhibition	0.305
10 d *D.magna* reproduction - juvenile production	-0.232
10 d *D.magna* reproduction - lethality	-0.555
Laboratory-based sediment bioassays	
10 d *C.riparius* lethality	0.704
28 d *C.riparius* growth	0.025
28 d *C.riparius* development	0.379
28 d *C.riparius* lethality	0.146
In situ bioassays	
6 d *G.pulex* feeding rate	0.534
6 d *D.magna* lethality	0.752
10 d *C.riparius* lethality	0.405

Table 3.58 shows the combinations of two, three and four bioassays that explained the highest levels of variability in the macroinvertebrate data. Combinations involving laboratory- based sediment tests (10 d *C.riparius* lethality and 28 d *C.riparius* growth/development/lethality) and *in situ* bioassays (6 d *D.magna* lethality and 6 d *G.pulex* feeding rate) were in all cases the better predictors. As endpoints were added

levels of the variability of macroinvertebrate abundance explained by the bioassay data increased.

Table 3.58 Combinations of bioassay endpoints which explained the highest levels of variability in the macroinvertebrate abundance data

No. of endpoints	Explanatory variables	Spearman's rank correlation
2	10d *C.riparius* lethality and 6d *G.pulex* feeding rate (*in situ*)	0.824
3	10d *C.riparius* lethality, 28d *C. riparius* growth and 6d *D.magna* lethality (*in situ*)	0.907
4	10d *C.riparius* lethality, 28d *C. riparius* growth and development and 6d *D.magna* lethality (*in situ*)	0.960

Tees Estuary

Table 3.59 summarises results from the Spearman's rank correlations between the macroinvertebrate abundance and bioassay data. The strongest single indicator of natural macroinvertebrate abundance in the Tees estuary was lethality of *C.volutator* in the laboratory over 10 days (Spearman's rank correlation=0.652). Other predictors with high Spearman's rank correlations were 28 d *Corophium* lethality tests in the laboratory, *in situ* 10 d *Corophium* lethality tests, cast formation of *A.marina* and Microtox. The ECL, 24 h OEL, 9 d *T.battagliai* reproduction and *in situ* 28 d *M.edulis* clearance rate bioassays were poorer predictors of macroinvertebrate abundance patterns. However, with the exception of *M.edulis* clearance rate test, all of these bioassays identified Station 5 as the most impacted.

Table 3.59 Spearman's rank correlations between macroinvertebrate abundance and bioassay results for the Tees estuary (ND = no data)

Explanatory variable	Spearman's rank correlation
Laboratory-based water column bioassays	
Enhanced chemiluminescence (ECL)	0.071
Bacterial bioluminescence (Microtox)	0.576
24 h Oyster embryo-larval development	0.349
9 d *T.battagliai* reproduction - juvenile production	0.425
Laboratory-based sediment bioassays	
10 d *C.volutator* lethality	0.652
28 d *C.volutator* growth	ND
28 d *C.volutator* lethality	0.309
10 d *A.marina* cast formation	0.411
In situ bioassays	
10 d *C.volutator* lethality	0.309
28 d *M.edulis* clearance rate	-0.779

Table 3.60 shows the combinations of two and three bioassays that explained the highest levels of variability in the macroinvertebrate data. Combinations involving laboratory-based sediment tests (10 d *C.volutator* lethality and 28 d *C.volutator* growth/development/lethality), the laboratory-based water column Microtox bioassay and the 10 d *C.volutator* lethality (*in situ*) bioassay were in all cases the better predictors. As endpoints were added increased levels of the variability of macro-invertebrate abundance were explained.

Table 3.60 Combinations of bioassay endpoints which explained the highest levels variability in the macroinvertebrate abundance data

No. of endpoints	Explanatory variables	Spearman's rank correlation
2	28d *C.volutator* lethality and Microtox	0.920
	10d *C.volutator* lethality (*in situ*) and Microtox	
3	28d *C.volutator* lethality, 10d *A.marina* cast formation and Microtox	0.943

Summary

The results from the River Aire suggest that a decline in water quality could be measured through monitoring of either macroinvertebrate richness and abundance, or the responses of freshwater bioassays. Unsurprisingly, bioassays with invertebrates were the best predictors of measured effects on natural macroinvertebrate assemblages in the River Aire. *In situ* water column crustacean tests and laboratory sediment bioassays using insect species achieved a high level of predictive power, with the 6 d *D.magna* lethality (*in situ*) bioassay and the 10 d *C.riparius* lethality laboratory sediment bioassays showing Spearman's rank correlations of 0.752 and 0.704 respectively. The ability to explain the variability in macroinvertebrate abundance data increased as additional endpoints were included in the assessment battery, such that a combination of four endpoints could explain 96% of the variability. It should be remembered that although the 7 d *L.minor* growth inhibition test was not strongly correlated with the macroinvertebrate survey data the results of the bioassay may can be indicative of eutrophication effects, which is valuable complementary information.

The greatest toxicity in bioassays and effects on natural populations of macroinvertebrates, were observed at Station 5 (the positive control) in the Tees estuary. As at the River Aire, a crustacean test was the best predictor of macroinvertebrate assemblage structure. However, in the Tees it was the sediment-dwelling amphipod *Corophium* volutator (in the 10 d sediment lethality bioassay) rather than the water column dwelling copepod *Tisbe battagliai* that was the best predictor. All of the tests, except the *Mytilus* edulis clearance rate (*in situ*) bioassay, were able to identify greater toxicity at Station 5 (the positive control).

The results obtained in the multivariate analysis for the River Aire and Tees Estuary sites are consistent with those obtained in studies from highly impacted locations in the United States. In these studies assessments of the macroinvertebrate communities have been compared with results obtained from bioassays of receiving waters samples themselves (Birge *et al* 1989, Clements and Kiffney 1994) and have generally

indicated that bioassay data can be predictive of macroinvertebrate survey data. Consequently it would appear that at locations where there is clear evidence of strong contaminant-induced impact the conclusions drawn from the results of a battery of bioassays and macroinvertebrate surveys may be comparable. However, this will not necessarily be the case for moderately impacted locations where the bioassays can detect subtle or intermittent effects which may or may not become apparent as changes in macroinvertebrate abundance and/or diversity.

3.5.3 Relationship of analytical chemistry data to ecotoxicological bioassay and biological survey data

No analytical chemistry data was collected as part of the study programme, therefore an assessment of the comparison between biological data (ecotoxicological bioassay responses and macroinvertebrate survey information) and analytical chemistry data can only be made using routine information collected by the local regulators during the period of the study (see Section 2.3.3).

River Aire

At the River Aire routine survey data for all the stations used in the study was only available for a limited number of determinands, namely; pH, dissolved oxygen, BOD , hardness, suspended solids unionised ammonia, , dissolved copper and total zinc (see Table 3.61). At Stations 2, 4 and 5 information was available on a karge number of determinands but this information could not be used in the multivariate analysis because there was no corresponding data for Stations 1 and 8.

Table 3.61 Summary of routine analytical chemistry data for receiving water samples taken at the River Aire stations during the study

Determinand	Concentration ranges at different stations during the study				
	Station 1	Station 2	Station 4	Station 5	Station 8
pH	8.30-8.44	7.63-8.05	7.55-7.77	7.39-7.75	7.44-7.59
Dissolved oxygen (%ASV)	72 - 92	62 - 68	71 - 82	68 - 80	51 - 59
BOD (mg l^{-1})	2.0 - 2.3	1.5 - 2.0	2.0 - 2.8	2.8 - 5.8	3.1 - 3.7
Hardness (mg $CaC0_3$ l^{-1})	168-187	178-186	146-159	153-165	206-228
Suspended solids (mg l^{-1})	2 - 3	2 - 6	2 - 7	6 - 17	11 - 13
Unionised ammonia (mg l^{-1})	1.6 - 5.2	0.6 - 2.5	1.4 - 2.1	1.0 - 2.3	12.3- 12.6
Dissolved copper (µg l^{-1})	2.0 - 5.7	2.8 - 3.9	3.8 - 5.5	5.3 - 5.6	5.2 - 7.2
Total zinc (µg l^{-1})	6 - 10	8 - 10	10 - 25	22 - 31	26 - 38

At the River Aire site the best chemical predictor of the pattern in the toxicity data was a single variable: the concentration of dissolved copper (r=0.732). However, since data was only available for a few toxicants at all stations (unionised ammonia, dissolved copper, total zinc) it is unlikely that copper was the single causative agent of measured toxicity.

Table 3.62 summarises results from the Spearman's rank correlations between the macroinvertebrate abundance and water chemistry data for the River Aire. The heavy

metals zinc and copper, plus pH, were best able to explain the pattern of macroinvertebrate abundance in the Aire. The sanitary determinands were poorer predictors of macroinvertebrate abundance patterns. However, since data was only available for a limited number of determinands it is unlikely that these parameters were the only causative agents of measured biological impact. The data does indicate that key physico-chemical parameters such as dissolved oxygen and hardness are not apparently implicated in the decline in biological quality along the River Aire.

Table 3.62 Spearman's rank correlations between macroinvertebrate abundance and measured water chemistry in the River Aire

Explanatory variable(s)	Spearman's rank correlation
pH	0.813
pH and zinc	0.907
pH and copper	0.841
pH, zinc and copper	0.876

Tees Estuary

At the Tees Estuary information was available on a wider range of chemical determinands than at the River Aire, namely sediment concentrations of trace metals (cadmium, chromium, copper, lead, mercury, nickel, zinc) and organic substances (dieldrin, lindane, hexachlorobenzene and PCBs) (see Table 3.63).

Table 3.63 Summary of routine analytical chemistry data for sediment samples taken at the Tees Estuary stations during the study

Determinand	Sediment concentrations at different stations (ug g^{-1})					
	1	2	4	5	7	8
Cadmium	1.7	1.3	1.4	1.1	0.4	0.5
Chromium	493	2143	215	81	61	78
Copper	339	190	144	154	49	53
Lead	465	202	180	85	63	86
Mercury	5.4	2.8	2.6	1.5	0.3	0.4
Nickel	49	47	39	57	30	36
Zinc	739	366	391	390	165	165
Dieldrin	3.8	<0.9	<0.9	1.6	<0.9	<1.3
Hexachlorobenzene	12.6	1.6	<2.3	100	<0.9	5.7
Lindane	<1.5	7.1	<0.9	1.9	<0.9	4.2
PCBs	93.8	37.9	36	57	34	51

Table 3.64 summarises the strongest sediment physico-chemical predictors of bioassay responses from the Tees. The heavy metals zinc and nickel, plus hexachlorobenzene, completely explained the observed pattern of toxicity in the bioassays. The highest levels of nickel and hexachlorobenzene were measured at Station 5.

Table 3.64 Spearman's rank correlations between bioassay and sediment physico-chemical data from the Tees estuary

Explanatory variable(s)	Spearman's rank correlation
Zinc	0.840
Nickel and zinc	0.956
Nickel and hexachorobenzene (HCB)	0.943
Nickel, zinc and hexachorobenzene (HCB)	1.00

Table 3.65 summarises results from the Spearman's rank correlations between the macroinvertebrate abundance and sediment physico-chemical data. Heavy metals, such as nickel, lead, zinc and cadmium, were best able to explain the pattern of macroinvertebrate abundance in the Tees. The measured organic chemical determinands were poorer predictors of macroinvertebrate abundance patterns. These results are generally consistent with those for the bioassay data (see Table 3.64).

Table 3.65 Spearman's rank correlations between macroinvertebrate abundance and sediment physico-chemical data from the Tees estuary

Explanatory variable(s)	Spearman's rank correlation
Nickel	0.652
Lead and zinc	0.880
Nickel and zinc	0.798
Cadmium, nickel and zinc	0.956
Lead, nickel and zinc	0.884

3.5.4 Summary

This study has shown that certain of the laboratory-based and *in situ* bioassays of receiving water column and sediment toxicity are predictive of patterns of macroinvertebrate assemblage structure at the highly impacted River Aire and Tees Estuary sites. Although a large number of bioassays was used in both the freshwater and estuarine studies, a smaller subset of assays was effective in predicting natural macroinvertebrate patterns of distribution and abundance. These findings have important implications for estuarine locations where there is no standardised procedure for carrying out macroinvertebrate surveys and the use of an appropriate battery of bioassays could represent a cost-effective means of monitoring receiving water quality.

Heavy metal (copper and zinc) concentrations were associated with toxicity and changes in macroinvertebrate assemblages at the River Aire, although they may simply have correlated with other, unmeasured toxicants. There was no evidence that sanitary determinands were responsible for the patterns of toxicity and assemblage structure measured in the River Aire.

The results from the Tees Estuary suggested that heavy metals in the sediments, principally zinc and nickel plus lead and cadmium, were associated with observed patterns in toxicity and macroinvertebrate assemblages. There was also a suggestion that HCB concentrations were associated with bioassay responses.

3.6 Summary of the suitability of bioassays for receiving water monitoring

3.6.1 Introduction

The assessment of the sensitivity, discrimination, power and ecological relevance of the bioassays used at the freshwater and estuarine sites in Sections 3.1 to 3.5 means that certain conclusions regarding their suitability for receiving water monitoring can be made. The data generated has been used to separate the methods into three groups:

1. those which appear suitable in their current form (or with minor modifications) for use in receiving water monitoring;

2. those which are potentially suitable for receiving water monitoring but which require additional development to address issues of interpretation or robustness or the preparation of a standardised procedure;

3. those which because of issues of sensitivity, discrimination or ecological relevance are not appropriate for receiving water monitoring.

Table 3.66 summarises the status of the different bioassays in the categories of laboratory-based water column and sediment tests and *in situ* tests and indicates where developments are needed. It is apparent from Table 3.66 that the following bioassays in each category are most appropriate for current use (in some instances with minor modifications):

1. *Laboratory-based water column tests*: 10 d *D.magna* reproduction test for freshwaters and the 24 h oyster embryo-larval development or 9 d *T.battagliai* reproduction test for estuarine waters;

2. *Laboratory-based sediment tests:* 10 d *C.riparius* lethality test for freshwater sediments and the 10 d *C.volutator* lethality test or 10 d *A.marina* cast formation for estuarine sediments;

3. *In situ tests:* 6 d *G.pulex* feeding rate test for freshwaters but there is no appropriate method for estuarine waters.

The data for the most suitable bioassays in each grouping are discussed in the following sections.

3.6.2 Laboratory-based water column tests

The combined juvenile production endpoints of the 10 d *Daphnia magna* reproduction test (a shortened first brood version of the OECD/ISO 21 day reproduction test) discriminated between the reference stations and other impacted stations at both freshwater sites used in this study (see Tables 3.43 and 3.44). At the River Aire site the lethality endpoint of the bioassay generally showed low effects (<10%) at the 'clean' upstream reference station (BMWP=149, ASPT=5.52) but marked responses (>80% lethality) at 3 of the 5 downstream stations of poor/bad biological quality

(where BMWP score were 22-47 and ASPT scores= 2.8-3.4) (see Figure 3.7). No differences in juvenile production were evident at stations showing low lethality. On the Willow Brook no marked lethality (<10%) was evident in any samples from any station but juvenile production data discriminated Station 1 (BMWP=19, ASPT=3.17) from the upstream reference station (BMWP=136, ASPT=4.86) (see Figures 3.9 and 3.10). A 48 h *D.magna* lethality test was not sufficiently sensitive at either site to discriminate between stations.

The approach adopted in the study meant the juvenile production endpoint had limited discrimination (see Table 3.49). However, subsequent studies of the shortened reproduction test have shown that extending the duration of the bioassay from 10 to 14 days and using 10 animals per sample replicate markedly improves the discrimination and power of the juvenile production endpoint by increasing the number of offspring produced and reducing within group variability (Environment Agency 1999a). Alternatively the full 21 day *D.magna* reproduction test could be used, although this would incur higher costs than the shortened 14 day version.

In the estuarine environment both the 24 h oyster embryo-larval development test and the 9 d *Tisbe battagliai* reproduction/lethality test both discriminated between the reference stations on the Tees and Mersey estuaries (Stations 8 and 1 respectively) and impacted stations (see Tables 3.45 and 3.46). At the clean reference stations on the Tees and Mersey estuaries Percent Net Response values in the OEL test and percentage lethality in the *T.battagliai* reproduction test were generally <10% while at stations shown by biological surveys to be impacted PNR values and percentage lethality were 100% (see Figures 3.18-3.23). The OEL test was shown to be capable of discriminating between stations showing a $\geq$5% difference in response. The lethality endpoint of the *T.battagliai* reproduction test was capable of discriminating between stations showing a $\geq$ 25% difference in response. The juvenile production endpoint of the *T.battagliai* reproduction bioassay (as for the same endpoint in the *D.magna* reproduction test) needed modification to improve the discrimination.

There were statistically significant correlations between the results of the OEL and *T.battagliai* reproduction tests at the two sites (see Table 3.36) indicating that there would be a degree of redundancy in using both bioassays in a battery to monitor receiving water quality. The 24 h OEL test is already used in the National Marine Monitoring Programme to assess the quality of estuarine and marine receiving waters and represents a more cost-effective option than the 9 d *Tisbe battagliai* reproduction test. Staff time to carry out the OEL test and generate the resulting data for a 40 sample study (8 station by 5 samples per station) is approximately 20 hours while this would be in the region of 50-55 hours for the 9 d *T.battagliai* reproduction test. However, the OEL test often produces an all or nothing response whereas the combined endpoints of the *T.battagliai* reproduction test offer greater discrimination between samples. Indeed in this study sub-lethal effects, in terms of reduced juvenile production, were evident in samples from the reference locations on the Tees and Mersey compared to numbers of offspring measured in the laboratory using a reference seawater (see Table 3.19). Samples from the Tees reference station (Station 8) showed a mean juvenile production of 8-33% of the levels in the reference seawater control while at the Mersey reference station (Station 1) numbers of offspring were 9-16% of those in reference seawater control. In contrast, the PNR values in the OEL

test were low at both reference stations on all occasions (see Figures 3.18 and 3.19). The ability of the 9 d *T.battagliai* reproduction test to identify low level effects at these reference stations is important because they were apparent before there was evidence of impact on macro-invertebrate abundance and biomass at these stations.

The rapid enhanced chemiluminescence (ECL) bioassay can be a potentially useful indicator of receiving water quality and discriminated between reference and impacted stations at all the study sites. At both freshwater and estuarine sites there were statistically significant correlations between ECL results and those from the 10 d *D.magna* reproduction test (lethality endpoint at River Aire site, see Table 3.35) at freshwater sites and the 24 h OEL test and 9 d *T.battagliai* reproduction tests (juvenile production endpoint at Tees Estuary and lethality endpoint at Mersey Estuary, see Table 3.36) at the estuarine sites. However, there is an issue regarding the interpretation of the ECL bioassay results. A review of the method has indicated that the bioassay responses are often strongly correlated with BOD/COD values for the samples tested and that the bioassay is responding to both toxicants and substances which are toxicologically innocuous to higher organisms (Environment Agency 1997b, 1999b). This issue needs to be resolved before the test can be used as a measure of toxicity only in receiving water samples.

However, individual laboratory-based water column bioassays were poor predictors of the variability in macroinvertebrate abundance data at the relevant location showing low Spearman's rank correlations (see Table 3.57).

3.6.3 Laboratory sediment bioassays

At the freshwater sites the 10 d *C.riparius* lethality bioassay was able to discriminate between the upstream reference stations and impacted downstream stations (see Tables 3.43 and 3.44). The test was shown to be capable of discriminating between stations showing a ≥13% difference in response (see Table 3.49). Furthermore, at the River Aire reference station (Station 1) the *C.riparius* lethality bioassay was able to detect the effects of an intermittent discharge. The 10 d *C.riparius* lethality bioassay was also shown to be a good predictor of macroinvertebrate abundance data on the River Aire (Spearman's rank correlation=0.704, see Table 3.57). There is a standardised procedure for the bioassay (ASTM 1997) and the test has been advocated for testing freshwater sediment quality in a SETAC 'Guidance Document on Sediment Toxicity Tests and Bioassays for Freshwater and Marine Environments' Hill *et al* 1993).

At the estuarine sites both the 10 d *C.volutator* lethality and 10 d *A.marina* cast formation were able to discriminate between the reference control stations and other stations (see Tables 3.45 and 3.46). At the Tees site both bioassays showed limited responses at the reference stations (where lethality was < 10% and the number of casts produced was >15) and marked effects at other stations (where complete lethality was observed and the number of casts was <5). At the Mersey Estuary site the 10 d *C.volutator* lethality bioassay showed lower levels of response compared to values measured at the Tees Estuary providing further evidence of the discriminatory capacity of the bioassay. The 10 d *C.volutator* lethality test was shown to be capable of discriminating between stations showing a ≥22% difference in response (see

Table 3.50). The design of the *A.marina* cast formation needed modification to improve test discrimination.

There was a statistically significant correlation between the results of the two tests at the Tees Estuary sites (see Table 3.38) indicating that there would be a degree of redundancy in using both in a battery to monitor receiving water quality. In terms of cost the staff time involved in conducting the 10 d *C.volutator* and 10 d *A.marina* cast formation bioassays is similar (approximately 80 h for the former against 70 h for the latter). The 10 d *C.volutator* lethality bioassay showed the highest degree of correlation with macroinvertebrate abundance data (Spearman's rank correlation=0.652).

The 10d *C.volutator* lethality and 10 d *A.marina* cast formation bioassays were recommended for assessing estuarine/marine sediment quality by Hill *et al* (1993) and were recommended for evaluating dredged material prior to relocation by the Sediment Bioassay Task Team of the Group Co-ordinating Sea Disposal Monitoring as part of the Marine Pollution Monitoring Management Group (MAFF 1997). These tests have been routinely used by CEFAS to monitor estuarine and marine sediment quality around the United Kingdom (for example MAFF 1995). The potential of the *C.volutator* lethality test for monitoring estuarine sediment quality is evidenced by its use by Environment Agency North Eastern Region as part of their assessment of the quality of the Tees (see Table 3.21).

3.6.4 *In situ* bioassays

At the freshwater sites the combined endpoints of the *G.pulex* feeding rate/lethality bioassay were capable of discriminating between the upstream reference station and impacted downstream stations. On the River Aire feeding rates of >0.25 mg food/ mg animal/day were measured at the upstream reference station (BMWP=149) whereas values of <0.15 mg food/mg animal/day were measured at the impacted downstream stations where BMWP values were 22-47. On the River Aire the 6 d *G.pulex* feeding rate bioassay data showed a strong correlation with macroinvertebrate abundance data (Spearman's rank correlation=0.534).

At the Willow Brook site the highest feeding rates (>0.2 mg food/ mg animal/day) were at Stations 6 and 8 (BMWP = 83-136) and the lowest (<0.15 mg food/mg animal/day) at impacted Stations 1 and 2 (BMWP = 12-19).

The test has been previously shown to be capable of identifying changes in receiving water quality due to continuous or episodic point source or non-point source discharges (Crane *et al* 1993, 1995).

Table 3.66 Summary of the suitability of different bioassays for receiving water monitoring

Type of test	Suitability of bioassay for receiving water monitoring	
	Freshwater	Estuarine
Laboratory-based water column		
4 min Enhanced chemilumnescence (ECL)	Potentially suitable (cost-effective, sensitive and response discriminates between stations) though issues relating to interpretation of response need to be addressed	Potentially suitable (cost-effective, sensitive and response discriminates between stations) though issues relating to interpretation of response need to be addressed
15 min Bacterial bioluminescence	Not suitable (relatively insensitive with many samples showing stimulatory responses and limited discrimination between stations)	Not suitable (relatively insensitive with many samples showing stimulatory responses and limited discrimination between stations)
7 d *L.minor* growth inhibition	Not suitable (relatively insensitive with many samples showing stimulatory responses and limited discrimination between stations) but could be useful for monitoring eutrophication	-
48 h *D.magna* immobilisation	Not suitable (insensitive and no discrimination)	-
10 d *D.magna* reproduction	Suitable (combined juvenile production and lethality endpoints are sensitive and able to discriminate between stations on two levels) but discrimination and power of the test can be improved by modifying test design	-
24 h Oyster (*C.gigas*) embryo-larval development	-	Suitable (sensitive and discriminatory between stations) but often all or nothing response
48 h *T.battagliai* lethality	-	Potentially suitable (response sensitive to samples from impacted locations but limited discrimination)
9 d *T.battagliai* reproduction	-	Suitable (combined juvenile production and lethality endpoints are sensitive and able to discriminate between stations on two levels) but discrimination and power of the test can be improved by modifying test design
Laboratory-based sediment		
10 d *C.riparius* lethality	Suitable (sensitive, able to discriminate between stations and predictive of macroinvertebrate survey data)	

28 d *C.riparius* growth	Potentially suitable (although lethality and development endpoints are sensitive and discriminatory between stations, there are issues relating to the interpretation of the growth endpoint) and further work is needed to improve the reliability of the method	
10 d *C.volutator* lethality	-	Suitable (sensitive, able to discriminate between stations and predictive of macroinvertebrate survey data)
28 d *C.volutator* growth	-	Potentially useful (although lethality endpoint is sensitive and discriminatory between stations, there are issues relating to the interpretation of the growth endpoint) and further work is needed to improve the reliability of the method
10 d *A.marina* cast formation	-	Suitable (cast formation endpoint is sensitive and discriminates between stations but test design needs modification whereas lethality endpoint is relatively insensitive)
In situ tests		
6 d *G.pulex* feeding rate/lethality	Suitable (combined juvenile production and lethality endpoints are sensitive, able to discriminate between stations on two levels and predictive of macroinvertebrate survey data)	-
6 d *D.magna* lethality	Potentially suitable (sensitive and able to discriminate between stations) but modifications to test design are needed to reduce variability and improve test power	-
10 d *C.riparius* lethality	Potentially suitable (sensitive and able to discriminate between stations) but modifications to test design are needed to reduce variability and improve test power	-
10 d *C.volutator* lethality	-	Potentially suitable (sensitive and able to discriminate between stations) but modifications to test design are needed to reduce variability and improve test power
28 d *M.edulis* clearance rate	-	Not suitable (sensitive but poor discrimination between stations due to high variability)

4. APPLICATION OF ECOTOXICOLOGICAL BIOASSAYS TO RECEIVING WATER MONITORING

4.1 Introduction

For many years the monitoring of receiving water quality has been largely based on the collection of water samples (and, in limited instances, sediment samples) followed by analysis for specific determinands. The parameters routinely measured have been the sanitary determinands; dissolved oxygen, ammonia and biological oxygen demand (BOD) and the List I (and II) substances defined under the European Community Dangerous Substances Directive (see Dobbs and Zabel 1994, Cartwright and Lewis 1996). This analytical data, along with biological (macroinvertebrate and fishery) survey data in freshwaters, has been used to determine management actions as part of a rolling programme.

The chemical-specific approach has proved valuable and cost-effective in situations where a limited number of determinands need to be measured in receiving waters and judged against defined international or national environmental quality standards. However, there is now such a diverse range of substances present in receiving waters for which standards have not been derived that adopting a chemical-specific approach in such situations has a number of fundamental problems (Parr 1994), including:

1. there can be uncertainty over which substances should be measured;

2. the analytical procedures needed to measure these substances may not have been established;

3. if analytical data on the concentrations of these substances can be generated there may not be ecotoxicological data available so that the potential effects on receiving water organisms can be determined;

4. chemical concentrations in environmental samples are often not accurate predictors of biological and ecological effects because the percentage of the chemical that is biologically available may range from 0-100% of the total concentration depending on the chemical and environmental conditions (for example pH, hardness, salinity, level of suspended solids).

Overall, the potential danger of the chemical-specific approach is that analyses for a large number of substances believed to be causing problems at a location will be carried out at considerable expense, and while receiving water concentrations of these substances are identified, these values cannot be adequately interpreted to define water quality. Furthermore, even if monitoring is conducted against existing standards the chemical-specific approach takes no account of the potential interactions between substances and the cost of monitoring increases as an increasing number of determinands are measured.

The use of macro-invertebrate surveys in water quality management also has limitations (see Section 1), primarily that:

- ecotoxicological techniques can detect more subtle changes;

- the approach cannot be applied to estuarine/marine waters (as there is no established methodology), highly polluted locations (since no or limited discrimination is possible) or channelised stretches of rivers (since sampling is not possible).

These problems can be addressed by adopting an integrated tiered chemical/ecotoxicological approach in which biological effects are initially measured alongside those determinands required under existing legislation. The information from the ecotoxicological methods provide a direct measure of biological effects and can be used to identify impacted areas in a catchment. Subsequently more intensive, but targeted, chemical and ecotoxicological analysis, as part of a Local Environmental Impact Assessment, can be directed towards these areas to identify the causative agent(s) of the biological effects. This approach may involve the use of Toxicity Identification Evaluations (TIE) which have been used applied more widely to the identification of substances responsible for effluent toxicity (Heber *et al* 1996).

4.2 Requirements of bioassays to be used in different applications

The requirements of bioassays to be used in a particular application (operational role) will depend on the objectives of that application. The objective of a general quality assessment scheme using biological effects measures is to generate a broad overview of the general quality status of receiving waters in order to monitor temporal trends, and to identify those stations at a site which need further investigation (local environmental impact assessment) with a view to remedial action. The information generated would be integrated with that from existing and proposed schemes based on other measures of water quality such as biology, chemistry, nutrients and aesthetics.

In contrast, the objective of local environmental impact assessment is to determine whether reductions in biological quality can be directly attributed to the impact of one or more point source discharges thereby allowing more effective discharge control measures to be implemented.

Obviously different types of information are required to meet these objectives although the two are linked in that the information generated in general quality assessment can be used to identify stations at a site which require the conduct of a local environmental impact assessment. The required level of certainty in the accuracy and environmental relevance of the toxicological information is much higher for local environmental impact assessments where the consequences of inaccurate assessments may have either environmental or economic implications for both regulators and industry. For GQA purposes, the level of certainty achieved must be counter-balanced by the cost of performing a large number of bioassays at the number of stations which would be involved in a monitoring programme. Any reduced uncertainty associated with the limitations of wide-scale application may be outweighed by the complementary information provided by other biological and chemical classification schemes and the fact that the local environmental impact assessment will confirm the extent of the problem. The level of certainty for both applications can be determined by the number of bioassays selected, the frequency of sampling, and the additional parameters measured to aid in bioassay interpretation (see Section 2.2.6).

An additional consideration is the nature of the toxicant inputs in both situations. General quality is determined by all point source and non-point source (diffuse) inputs into the receiving system, while local impact is determined by the point source discharge(s) of interest. Single point source discharges for local environmental impact assessments may be well defined in terms of composition, volume, frequency of discharge, receiving water dilution, size of the mixing zone and in some cases toxicity. However, it is not always possible to define the nature, composition and frequency of all inputs contributing to the general quality of a stretch or area of receiving water. These differences in sources are important in defining the type of bioassays used for both applications.

4.2.1 Bioassays requirements for general quality assessment

As discussed above, the general quality of receiving waters is dependent on point and non-point (diffuse) sources of pollution integrated over time. This integration is reflected in the composition of biological communities that are resident at locations of interest which means that biological classification should be a relatively consistent measure of 'general quality' over time. However, the integrated ecotoxicological effect of all inputs into a system is difficult to measure in spot samples of water and sediment taken at one point in time. The influence of each input can vary widely on a small temporal scale. Point sources such as industrial, urban or sewage discharges may be continuous or intermittent depending on industrial and treatment processes and weather-related flow events. Non-point sources of pollution such as agricultural and urban run-off are often determined primarily by weather conditions such as the frequency and duration of rain events. Conditions in the receiving water which influence the dilution of toxicants from both point and non-point discharges are also variable depending on weather conditions. An additional factor influencing toxicant concentrations in estuarine systems is tidal cycle. Therefore, extreme variability of discharge, dilution and tidal conditions can lead to a high degree of temporal variability in measures of receiving water toxicity. This obviously has implications for general quality assessment and the question of temporal variability must be addressed if the biological effects measures are to accurately reflect 'general quality' over time (see Section 5).

Of the three types of bioassay used in this study, laboratory-based water column bioassays were the most temporally variable as toxicant concentrations in the water column can change at any one sampling point on very small temporal scales (see Section 3.1). Temporal variability was also demonstrated for the freshwater and estuarine laboratory sediment bioassays, though the extent of this was not as marked as for water column bioassays. The flow of toxicants over the sediment surface will obviously influence concentrations in the sediment porewater, and concentrations will also change depending on toxicant concentrations in station water if this is used in the sieving process. An element of spatial variability may also be responsible for the temporal variability observed in sediment bioassays as it is not possible to sample exactly the same sediment on two occasions due to small spatial differences in sediment type. In addition, disruptions of sediment integrity during sampling and handling may also contribute to differences in toxicity seen on different sampling occasions. However, because of their stationary nature, sediments should be better at integrating concentrations of hydrophobic substances from point and non-point sources over time.

Of the three types of test used in this study, the most effective in integrating time-related variability were the *in situ* bioassays (see Section 3.1). Although temporal variability can be

seen in these bioassay results, the difference in response between deployments is more likely to reflect the real differences in quality over extended durations. Because the organisms are exposed to receiving water conditions for a minimum of six days, the chances of exposure to intermittent pulses of toxicants is more likely than if a series of spot samples alone are tested.

Therefore, for general quality assessment it is recommended that a combination of complementary *in situ* water column and sediment bioassays are used for general quality assessment as part of an integrated strategy. Tests with combined lethal and sub-lethal endpoints are preferable because they can discriminate between stations at a site on two levels with stations exposed to high concentrations of contaminants showing lethal effects while sub-lethal effects can be measured at stations exposed to lower chemical contaminant concentrations.

This approach is consistent with the findings of the multivariate analysis of ecotoxicological and macroinvertebrate survey data for the River Aire and Tees Estuary sites where laboratory based sediment tests and *in-situ* bioassays were the most effective predictors of biological quality (see Section 3.5).

Table 4.1 summarises the mechanism by which the data from the *in situ* and sediment bioassays used in general quality assessment could be interpreted within the context of an integrated approach (see Table 1.1).

Table 4.1 Summary of the interpretation of *in situ* and sediment bioassays (⌋ - effect, x - no effect)

In situ bioassays response		Sediment bioassay	Conclusion	Action
Lethal	Sub-lethal	responses		
x	x	x	No toxicity issue	No action with respect to toxicity testing*
⌋	x	⌋ or x	Lethal toxicity in water column	Investigate sources and causes
x	⌋	⌋ or x	Sub-lethal toxicity in water column	Investigate sources and causes
x	x	⌋	No water column toxicity but sediment toxicity	Consider sediment remediation but no action on discharges

* - There may be chemical or biological survey issues

In terms of using the toxicity data generated within a toxicity-based classification scheme a number of approaches are possible including:

1. considering the results of each bioassay as a separate entity and classifying stations according to the pattern of effects shown;
2. combining the toxicity data from the bioassays into a hazard index.

If a hazard index is calculated the mean response could be expressed as a percentage effect compared to that at a reference site or the total percentage effect possible in the bioassays used. The use of the total possible percentage effect may be a more justifiable way of performing any classification so that the results are comparable across receiving water systems rather than being determined by the quality of the reference station at a site. As long as the same test battery is used at each station at different sites the use of the total possible percentage effect would ensure geographical comparability.

In order to be comparable with chemical and biological classification schemes for receiving water quality both approaches would need to be able to separate stations into a series of classes ranging from good to bad (see Introduction). An example of a classification scheme is given in Appendix A.

4.2.2 Bioassays requirements for local environmental impact assessments

Local environmental impact assessments are concerned with the adverse effects of one or more point-source discharges in the receiving water surrounding the discharge point(s). As discussed earlier, the nature of single point-source discharges tends to be better defined than the combination of point and non-point source discharges contributing to general quality. Therefore, for local environmental impact assessments a sampling strategy can be designed which takes into account worst-case conditions in the receiving water thus reducing the temporal variability associated with general quality. Worst-case conditions in the water-column and sediment in the vicinity of a point-source discharge will be found when concentrations of toxicants in the discharge are known to be at a maximum and dilution is minimised at periods of low flow or low tide. Perhaps of more interest in the impact assessment of point-source discharges is the relationship between the toxicity of the discharge and the receiving water biological effects. Determining this relationship enables the efficiency of any toxicity-based discharge limits to be assessed in terms of their ability to protect the receiving water environment. This can be best achieved by using a common toxicological currency for discharge and receiving water testing.

Because the tests for assessing the end of pipe toxicity of discharges are by nature laboratory-based water-column tests, these assays should also be given priority for receiving water impact assessments and the selected methods should be as similar as possible to those proposed for setting toxicity-based limits. However, in cases where point-source discharges contain hydrophobic substances which are likely to accumulate in sediment or high levels of particulate matter, the use of sediment toxicity bioassays should also be considered when monitoring receiving water quality.

Additional measures such as flow rate, temperature, ammonia and suspended solids should be measured in local environmental impact assessments in order that the effects of other stressors can be dissociated from toxicant effects and adverse effects measured in the receiving water can be linked back to the discharge of interest.

The sampling design selected for local impact assessment is highly site and discharge specific. Different combinations of bioassays may be chosen at each site and the number of sample points and sampling occasions should be decided on based on the perceived importance of temporal and spatial variability. Ultimately the selection of bioassays should be driven by the choice of tests for measurement of discharge toxicity and also a risk assessment of the potential impact of the discharge on the receiving water environment.

5. ISSUES RELATED TO THE USE OF BIOASSAYS IN A PARTICULAR OPERATIONAL ROLE

5.1 Introduction

In terms of the use of bioassays in a particular operational role there are a number of issues which need to be considered when designing both the sampling/deployment strategy (number of sampling/deployment stations at a site, number of samples taken) and the testing strategy (number of replicates of laboratory-based tests or cages used in *in situ* tests). These issues include:

1. consideration of the potential influence of temporal variability (see Section 4);

2. consideration of the potential influence of small scale spatial effects at a station;

3. consideration of the power required from a bioassay to be able to detect pre-defined differences in response between stations (see Section 3.4).

These issues are considered in the following sections with reference to the use of bioassays for general quality assessment and, where appropriate, local environmental impact assessments.

5.2 Temporal variability

Temporal variability was found in the results of data for most bioassays but this is not unexpected given the dynamic nature of the environments being monitored. Changes in response over time, which have also been found for macro-invertebrate survey data, reflect changes in:

- contaminant release from point sources and diffuse inputs;

- available dilution due to changes in river depth or tidal state in estuaries.

However, if it is deemed appropriate to overcome the issue of temporal variability two possible means of achieving this are to:

(i) perform bioassays as many times as possible over a specified sampling window, taking into account as many different weather conditions, flow rates and tidal cycles as possible. The mean response for all the sampling occasions could then be used as a basis for classification. The difficulty with this approach is that it is potentially labour intensive and the optimum number of sampling times is difficult to define.

(ii) perform the sampling and testing only once during a period defined *a priori* as 'worst-case'. This one measurement could then be used for classification purposes. This approach is less labour intensive but may lead to inaccurate classification particularly if worst-case conditions are not adequately defined.

In order to design a cost-effective approach to toxicity-based classification, aspects of both approaches can be combined. Sampling and testing could be performed on more than one occasion, but the sampling times are restricted so that only defined 'worst-case' conditions are considered. The number of sampling times is defined by logistical constraints but it is likely that a minimum of two occasions would be necessary. In order to reduce the influence of temporal variability further only those procedures which are likely to integrate the effects of short-term fluctuations in receiving water toxicant concentrations would be selected.

For freshwater systems two worst-case scenarios are possible. For non-point source discharges a worst-case scenario will be the first flush of a storm-event after a period of dry weather. This scenario would include any increase in toxicant concentrations as a result of agricultural and urban run-off and any flow-related point source discharges such as storm-sewer overflows. *In situ* bioassays could be deployed prior to the storm event to maximise exposure to the first flush. For point-source discharges, a worst case scenario will be a period of low flow when receiving water dilution is at a minimum. Again the *in situ* bioassays could be deployed during such a period. Sediment samples could be taken on both occasions at the end of the *in situ* deployments to increase the chances of measuring the effects of toxicants which may also have impacted the organisms exposed *in situ*. A sediment bioassay performed following the first flush would measure the effects of toxicants from both non-point and point sources but would also include any recent inputs following the storm event. A sediment test performed at low flow would measure the effects of toxicants from both point source and non-point sources at a period of low flow when the maximum response would be expected. The mean response of both the sediment and *in situ* bioassays measured during both the worst-case scenarios could then be used as a basis for classification.

Although *in situ* bioassays are thought to integrate changes in exposure over time this study demonstrated that the results can still be temporally variable and so the deployment times must be carefully selected on the basis of weather conditions and flow regime.

Sediment toxicity is generally thought to be less temporally variable than water toxicity due to the capacity of the former to integrate and accumulate contaminants over time. Although this may be true, the results of this study have shown that sediment toxicity can also be temporally variable. This is probably due to the fact that the same sediments cannot be sampled on different occasions and also that sediment porewater will contain concentrations of toxicants which have recently passed over the sediment surface. Careful consideration must be given to when sediment samples should be removed for toxicity assessment, based on weather conditions and flow regimes.

For estuarine systems, the general quality assessment need not be performed during a period of dry weather as weather-related non-point source inputs are less likely to be of concern. However, the state of the tide at which samples are taken needs to be carefully considered. As for the freshwater assessment, sediment bioassays could be performed at the end of the *in situ* deployment to increase the likelihood of measuring toxicants which may have impacted the organisms exposed *in situ*.

5.3 Small scale spatial variability

In the environment some small scale spatial variability may be expected at a station and for general quality assessments this needs to be considered in the design of the sampling/deployment programme. Although it is not necessary to have an exact estimate of the extent of small scale spatial variability for general quality assessments a representative estimate of the general level of toxicity at a station may not be obtained if either sufficient samples are not taken for laboratory-based bioassays or sufficient cages are not deployed for *in situ* bioassays. For laboratory-based assays compositing of samples may be adopted to reduce costs provided this approach still leads to a realistic measure of the general level of toxicity at a station. For local environmental impact assessment small scale spatial variability may be of greater interest and in this case it will be more appropriate to carry out laboratory-based bioassays on a series of samples collected at the station.

The statistical analyses of the data generated in this study (see Section 3.1) have indicated that generally laboratory-based water column and sediment bioassays and *in situ* bioassays showed no statistically significant small scale spatial effects (that is the samples were generally homogeneous). However, this was not the case for sediment samples measured with longer-term (28 d) sediment bioassays, which were apparently not homogeneous. This may be due to small differences in sediment type between samples taken from the same stations. It is possible that over a longer exposure duration the test organisms became more sensitive to substrate type and nutritive quality (Ingersoll 1995). The fact that the 28 d *C. riparius* and *C. volutator* growth lethality bioassay received additional food, did not seem to factor out the influence of sediment type. If sediment type was responsible for the spatial variability observed in the longer duration bioassays, it is likely that differences observed between stations were also likely to be influenced by differences in sediment type. In order to differentiate between the effects of sediment type and the effects of toxicant concentrations in this study, comparisons were made with previous studies where survival of the test organisms was observed in clean sediments with a range of physico-chemical properties. However, a more rigorous approach would be to construct regression models of sediment type against control response. This should be performed for more physico-chemical determinands than those measured in this study. For example, ammonia toxicity within sediment is likely to affect bioassay response. Given the wide range of sediment types likely to be found on a national level, this should be considered for any sediment test selected for classification, but is particularly important if chronic tests are used.

Table 5.1 shows the correlation coefficients (and significance) between bioassay responses in spot and composite water column and sediment samples from the case study sites. From the data it is apparent that for most of the laboratory bioassays the responses measured in the multiple spot samples were significantly correlated with those found in composite samples. The only exceptions were the Microtox-acute bioassay at the Willow Brook, the growth and lethality endpoints in the 28 day *C. riparius* bioassay at the River Aire and the growth endpoint in the 28 day *C. volutator* bioassay at the Mersey Estuary. Therefore, it would appear that the use of composite samples would provide a cost-effective means of monitoring a larger number of stations in a general quality assessment programme for receiving waters. However, it is important that the same number of randomly obtained samples are used to prepare the composites from each station to avoid any potential bias (ASTM 1998). Furthermore it may be

necessary to increase the number of test replicates of test organisms used in each bioassay to achieve an acceptable level of power in the tests.

5.4 Power of the bioassays to detect differences in response

The issue of the power of bioassays required for general quality assessment or local environmental impact assessment has been discussed in detail in Section 3.4. In terms of the level of power required in bioassays used for general quality assessment or local environmental impact assessment this could be defined in a pragmatic manner by simply requiring that a certain percentage change from the reference control level (for example 10 or 20%) can be identified. However, it would be more environmentally meaningful if the level of change required could be related to population parameters such that if the change in bioassay response at a station was detected this would have clear implications for population dynamics. For example, Crane *et al* (1993) found that a 25-30% difference in feeding rate in the *G. pulex in situ* bioassay was consistent with changes in individual growth and reproductive output. Studies in the Netherlands on *D. magna* population dynamics have shown that 25-30% changes in reproductive output over 21 days can, for substances such as 3,4-DCA, result in a 20% reduction in the intrinsic rate of population increase, r_m (TNO 1991).

Table 5.1 Correlation coefficients (and significance) between bioassay responses in spot and composite water column and sediment samples from the case study sites

Bioassay	Site	Correlation coefficient (and significance) between bioassay responses in spot and composite samples
Freshwater		
ECL	River Aire	0.942 ($p<0.001$)
	Willow Brook	0.829 ($p<0.001$)
Microtox-acute	River Aire	0.341 (NS)
	Willow Brook	0.947 ($p<0.001$)
L. minor growth inhibition	River Aire	0.927 ($p<0.001$)
	Willow Brook	0.946 ($p<0.001$)
D. magna reproduction		
- juvenile production	River Aire	0.875 ($p<0.001$)
	Willow Brook	0.876 ($p<0.001$)
- lethality	River Aire	Not calculable
C. riparius lethality (10 d)	River Aire	0.925 ($p<0.001$)
	Willow Brook	0.853 ($p<0.001$)
C. riparius growth (28 d)	River Aire	-0.039 (NS)
C. riparius development (28 d)	River Aire	0.934 (($p<0.01$)
C. riparius lethality (28 d)	River Aire	0.735 (NS)
Estuarine		
ECL	Tees Estuary	0.992 ($p<0.001$)
	Mersey Estuary	0.990 ($p<0.001$)
Microtox-acute	Tees Estuary	0.990 ($p<0.001$)
	Mersey Estuary	0.930 ($p<0.001$)
OEL	Tees Estuary	0.996 ($p<0.001$)
	Mersey Estuary	0.980 ($p<0.001$)
T. battagliai reproduction		
- juvenile production	Tees Estuary	0.685 ($p<0.001$)
	Mersey Estuary	0.760 ($p<0.001$)
- lethality	Tees Estuary	0.860 ($p<0.001$)
	Mersey Estuary	0.867 ($p<0.001$)
C. volutator lethality (10 d)	Tees Estuary	0.922 ($p<0.001$)
	Mersey Estuary	0.602 ($p<0.05$)
C. volutator growth (28 d)	Tees Estuary	0.990 ($p<0.001$)
	Mersey Estuary	0.07 (NS)
C. volutator lethality (28 d)	Tees Estuary	0.961 ($p<0.001$)
	Mersey Estuary	0.871 ($p<0.01$)
A. marina cast formation (10 d)	Tees Estuary	0.931 ($p<0.001$)
A. marina lethality (10 d)	Tees Estuary	0.659 ($p<0.05$)

NS - Not significant

6. CONCLUSIONS

The testing of field collected water and sediment samples with laboratory-based water column and sediment bioassays and the deployment of *in situ* bioassays at two riverine (River Aire and Willow Brook) and two estuarine (Tees and Mersey estuaries) sites in a summer window (July to October 96) resulted in the following conclusions being drawn:

- on the basis of the sensitivity, discrimination between stations and statistical power of the bioassays a number of methods of each type can be used (in there present format or with minor modifications) to assess receiving water quality:

Type of bioassay	Suitable candidates for different environments	
	Freshwater	Estuarine
Laboratory-based water column	10 d *D.magna* reproduction	9 d *T.battagliai* reproduction or OEL
Laboratory-based sediment	10 d *C.riparius* lethality	10 d *C.volutator* lethality or 10 d *A.marina* cast formation
In situ	6 d *G.pulex* feeding rate	None currently appropriate

The bioassays given in the table above all showed a range of responses from limited effects at reference stations to marked effects at impacted stations. They were able to discriminate statistically between reference and impacted stations since in many instances the bioassays were able to separate stations on the basis of differences in response of 25% effect or lower. A number of the bioassays had both sub-lethal and lethal endpoints which enabled stations to be separated on two levels:

- comparisons of the bioassay data and macroinvertebrate survey data on the River Aire and Tees Estuary have shown that certain of the laboratory-based and *in situ* bioassays of receiving water column and sediment toxicity are predictive of patterns of macroinvertebrate assemblage structure at these highly impacted sites. Although a large number of bioassays was used in both the freshwater and estuarine studies, a smaller subset of assays was effective in predicting natural macroinvertebrate patterns of distribution and abundance. These findings have important implications for estuarine locations where there is no standardised procedure for carrying out macroinvertebrate surveys and the use of an appropriate battery of bioassays could represent a cost-effective means of monitoring receiving water quality. However, certain bioassays were also able to detect the effects of intermittent discharges or sub-lethal effects at reference locations which had not resulted in impacts on the macroinvertebrate community;

- a number of longer-term laboratory-based sediment bioassays (28 d *C.riparius* growth and 28 d *C.volutator* growth) or *in-situ* bioassays (6 d *D.magna* lethality and 10 d *C.volutator* lethality) could be potentially useful for receiving water monitoring but further standardisation is needed before they can be used routinely.

- the type of bioassays used for a particular operational role depends on the objectives of the study. For example in the case of general quality assessment where monitoring has to consider the effects of intermittent and continuous point and non-point sources the use of laboratory-based sediment bioassays and *in situ* bioassays are considered to be the most appropriate. For local environmental impact assessments of the impact of specific discharges laboratory-based water column bioassays are considered to be most appropriate.

- the use of composite water column or sediment samples represents a means of reducing the costs associated with general quality assessments whilst still accounting for small scale spatial effects at a locations.

REFERENCES

Akberali, H.B. and Trueman, E.R. (1985) Effects of environmental stress on marine bivalve molluscs. *Adv. Mar. Biol.* **22**, 102-199.

Baldwin, I.G. and Kramer, K.J.M (1994) Biological Early Systems (BEWS). In *Biomonitoring of coastal waters and estuaries* edited by K.J.M Kramer, CRC Press, Inc, Boca Raton, 1-29.

Bayne, B.L., Widdows, J. and Thompson, R.J. (1976) Physiology II. In *Marine Mussels: their Ecology and Physiology,* edited by B.L Bayne, Cambridge University Press, 207-260.

CEFAS (1997) Final Report of the Sediment Bioassay Task Team of the Marine Pollution Monitoring Management Group of the Group Co-ordinating Sea Disposal Monitoring. Aquatic Environment Monitoring Report No. 48. Lowestoft.

Coleman, N. (1974) The heart rate and activity of bivalve molluscs in their natural habitat. *Oceanogr. Mar. Biol. Ass. UK* **32**, 175-192.

Crane, M., Maltby, M. and Johnson, I. (1993) The Development of the *Gammarus pulex* feeding rate assay. Final Report to the National Rivers Authority, R&D Note 170.

Depledge, M.H. (1989) The rational basis for the detection of the early effects of marine pollution using physiological indicators. *Ambio* **18**, 301-302.

Depledge, M.H. and Anderson, B.B. (1990) A computer-aided physiological monitoring system for continuous, long-term recording of cardiac activity in selected invertebrates. *Comp. Biochem. Physiol* **96A**, 474-477.

Depledge, M.H. (1991) Conceptual paradigms in marine ecotoxicity. In: *Proceedings of the 12th Baltic Marine Biologists Symposium,* edited by E. Bjornstad, L. Hagerman and K. Jensen, 47-52, Olsen and Olsen, Fredensborg, Denmark.

Depledge, M.H. (1993) The rational basis for the use of biomarkers as ecotoxicological tools. In: *Non-destructive biomarkers in vertebrates*, edited by M.C. Fossi and C.L. Leonzio, 261-285. Lewis Publishers, Boca Raton, FL, USA.

Depledge, M.H. (1994) The rational basis for the use of biomarkers as ecotoxicological tools. In: *Non-destructive biomarkers in vertebrates*, edited by M.C. Fossi and C.L. Leonzio, 261-285. Lewis Publishers, Boca Raton..

Depledge, M.H. and Lundebye, A.K. (1996) Physiological monitoring of contaminant effects in individual rock crabs, *Hemigrapsus edwardsi*: The ecotoxicological significance of variability in response. *Comp. Biochem. Physiol*, **113C**: 277-282.

DeWitt, T.H., Ditsworth, G.R. and Swartz, R.C. (1988) Effects of natural sediment features on survival of the phoxocephalid amphipod, *Rhepoxynius abronius, Marine Environmental Research,* **25:** 99-124.

Dieter, H.H., Muller, L., Abel, J. and Summer, K.H. (1987) Metallothionein-determination in biological materials: Inter-laboratory comparison of 5 current methods. In: *Metallothionein II*, edited by J.H.R Kägi and Y. Kojima. Birkhäuser Verlag, Basel, 52-351.

Environment Agency (1997a) Toxicity-Based Criteria for Receiving Water Quality (Stage 1). R&D Technical Report P26.

Environment Agency (1997b) Review of the Enhanced Chemiluminescence (ECL) Test. R&D Technical Report.

Fleming, R., Vine, S., Hermsen, W., Spence, S. and Crane, NM. (1993) Partitioning of poorly-soluble substances in sediments. Final Report to the National Rivers Authority, Project Leader: Roger Milne.

Fleming, R. and Nixon, S. (1996) Sediment Toxicity Tests. Final Report to the DoE, April 1996.

Fleming, R., van de Guchte, C., Grootelaar, L., Ciarelli, S., Borchert, J., Looise, B. and Guerra, M.T. (1996) Sediment tests for poorly water soluble substances: European intra- and inter- laboratory comparisons. In: *Development and Progress in Sediment quality Assessment: Rationale, Challenges, Techniques and Strategies,* edited by M. Munwar and G. Dave, SPB Academic Publishing, Amsterdam, The Netherlands.

Fleming, R. and Sims, I. (1997) Development of sediment bioassays: Toxicity-Based Criteria, Stage 1, Draft Report to the Environment Agency, Project Leader: Nerina Holden.

Giddings, J.M. (1986) Protecting aquatic resources: an ecologists perspective. In: *Aquatic toxicology and environmental fate*, edited by T.M. Poston and R. Purdy, 97-106. ASTM STP 921 American Society for Testing and Materials, Philadelphia, P.A., USA.

Grace, A.L. and Gainey, L.F. (1987) The effects of copper on the heart rate and filtration rate of *Mytilus edulis*. *Mar. Poll. Bill* **18**, 87-91.

Hogstrand, C,. and Haux, C. (1990) Metallothionein as an indicator of heavy metal exposure in two sub-tropical fish species. *J. Exp. Mar. Biol.* **138**, 69-84.

Holmes, D., Delaney, P. and Fleming, R. (1994) Toxicity of iron in sediments. WRc Technical Report for Anglian Water Services, Draft.

Kohler, A. (1989) Cellular effects of environmental contamination in fish from the River Elbe and North Sea. *Mar. Env. Res.* **28**, 417-424.

Lowe. D.M., Fossato, V.U., Moore, M.N. (1993) Lysosomal membrane damage as an *in vitro* marker of contaminant impact under field and laboratory conditions. I.C.E.S. Marine Environmental Quality Committee. C.M. 1993/E:32.

Lowe, D.M., Pipe, R.K. (1994) Contaminant induced lysosomal membrane damage in marine mussel digestive: an *in vitro* study. *Aquat. Toxicol.* **30**, 357-365.

Mayer, F.L., Versteeg, D.J., McKee, M.J., Folmar, L.C., Graney, R.L., McCume, D.C., Ratner, B.A. (19<>) Physiological and non-specific biomarkers. In: *Biomarkers; Biochemical Physiological and Histological Markers of Anthropogenic Stress*, edited by R.J. Hugert, R.A. Kimerle, P.M. Mehrle, H.L. Bergman, 5-85. Lewis Publishers, Boca Raton..

Moore, M.N. (1985) Cellular responses to pollutants. *Mar. Pollut. Bull.* **16,** 134-139.

Moore, M.N., Webberburn, R.J., Lowe, D.M. and Depledge, M.H. (1996) Lysosomal reaction to xenobiotics in mussel hemocytes using BODIPY-FL-verapamil. *Mar. Env. Res.* Vol. **42,** No. 1-4, 99-105.

Pederson, K.L., Pederson, S.N., Højrub, P., Andersen, J.S., Roepstorff, P., Knudsen, J., and Depledge, M.H. (1994) Purification and characterisation of a cadmium-induced metallothionein from the shore crab *Carcinus meanas* (L). *Biochem. J.* **197**, 609-614.

Roch, M. McCarter, J.A., Matheson, A.T., Clark, M.J. and Olafson, R.W. (1982) Hepatic metallothionein in rainbow trout (*Salmon gairdneri*) as an indicator of metal pollution in the Campbell River system. *Can. J. Fish. Aquat. Sci.* **39**, 1597-1601.

Sabourin, T.D. and Tullis, R.E. (1981) Effects of three hydrocarbons on the respiration and heart rates of the mussel, *Mytilus californianus. Bull. Environ. Contam. Toxicol.* **26**, 729-736.

Viarengo, A., Ponzano, E., Dell'Anno, A. and Fabbri, R. (1995) Simple methods for metallothionein quantification: an application to detect metallothionein concentration in the tissues of Mediterranean and Antarctic molluscs. In: *Ecotoxicology and the Marine Environment*, Vol. **2**, edited by P.D. Abel, American Company Paramount Communications.

Zar, J.H. (1984) Biostatistical Analysis, Second Edition, Prentice-Hall International, Inc.

APPENDIX A EXAMPLE OF A CLASSIFICATION SCHEME USING BIOASSAY DATA

The most pragmatic approach to deriving a toxicity-based classification scheme is to combine the results of all the selected bioassays into a hazard index. The hazard index can then be categorised into classes ranging from good to bad in order to maintain consistency with existing and proposed biological and chemical classification schemes (see Section 4.3.1)

To calculate the index, the mean response of each toxicity bioassay, performed on a minimum of two occasions, should be calculated. In order to ensure comparability between all bioassays and endpoints measured, the mean response should be expressed as a percentage effect. This could be the percentage effect compared to a reference site or simply the total percentage effect. The total percentage effect should be used where no reference sites were defined in the study design. This may be a more justifiable way of performing a classification so that the results are comparable across receiving water systems rather than being determined by the quality of the reference site within each system. As long as the same tests are used at each station in each system, total percentage effect would ensure geographical comparability. The bioassay data used in deriving the hazard index may comprise shorter ($\lceil$10 day) or longer term (28 day) information on sub-lethal (growth, development, feeding rate or reproduction) or lethal responses.

However, the use of both shorter-term and longer-term bioassays, and lethal and sublethal endpoints in a battery will mean that individual responses will carry different levels of importance in terms of defining the toxicity within a system.

For consistency all the data needs to be expressed as a percent effect. In the case of the lethality endpoint data can be used directly whereas for sub-lethal responses it will be necessary to express the given endpoint value as a percentage of that in an appropriate reference. No effect would be expressed as a score of 0 while 100% effect (% lethality or % reduction in feeding rate, juvenile production or development) would be expressed as a score of 1.0. For example, for a given freshwater location using hypothetical data:

30% lethality in the 10 d *C. riparius* sediment test = score of 0.3
50% reduction in the 6 d *G. pulex* (feeding rate) *in situ* test = score of 0.5.

Although shorter- and longer-term lethality will have the same ecological effect, the shorter-term bioassays demonstrate a higher degree of toxicity within the system. Therefore, these bioassays could be given a higher weighting. As the relationship between shorter- and longer-term effects is difficult to define, particularly in different mixtures of contaminants, the most pragmatic solution is to add a weighting factor of two to the shorter-term bioassays. The longer-term lethal and sublethal endpoints should be expressed without a weighting factor.

In order to perform a classification the maximum possible hazard index should be calculated for the results of the bioassay battery.

For the data given above the hazard index derived would be:

Hazard Index = 2 x (0.3) + 2 x (0.5) = 1.6

In order to delineate class boundaries the maximum possible hazard index for a combination of bioassays can be divided into categories on the basis of percentage response. One approach may be to divide the classes as follows:

Response	Class
<20% of maximum hazard index	Good
20-50%	Fair
50-80%	Poor
80-100%	Bad

This is based on the fact that <20% response is regarded in most bioassay protocols to indicate the range of control response (that is no adverse effect) and >50% is usually judged to be unacceptable. Definition of these class boundaries is relatively arbitrary and there will be areas of uncertainty between the classes particularly at stations of moderate toxicity. However, the ecological relevance of the boundaries can only be assessed by further comparisons with biological classification on a wider scale.

On the basis of the maximum hazard indices calculated for a freshwater battery of the laboratory based 10 d *C.riparius* lethality bioassay and the *in situ* 6 d *G.pulex* feeding rate, the class boundaries would be as follows:

% Response	Class	Freshwater* class boundaries (based on hazard index)
<20%	Good	<0.8
20-50%	Fair	0.8-2.0
50-80%	Poor	2.0-3.2
80-100%	Bad	3.2-4.0

* - based on a maximum hazard index of 4

The certainty associated with these class boundaries will increase as the maximum hazard index and the size of classes increases. This can only be achieved by increasing the number of bioassays used.