Environment Agency

Evidence

Standardisation of RIVPACS for deep rivers: Phase I - review of techniques for sampling benthic macro-invertebrates in deep rivers

The Environment Agency is the leading public body protecting and improving the environment in England and Wales.

It's our job to make sure that air, land and water are looked after by everyone in today's society, so that tomorrow's generations inherit a cleaner, healthier world.

Our work includes tackling flooding and pollution incidents, reducing industry's impacts on the environment, cleaning up rivers, coastal waters and contaminated land, and improving wildlife habitats.

This report is the result of research commissioned and funded by the Environment Agency's Science Programme.

Published by:
Freshwater Biological Association

August 2014

All rights reserved. This document may be reproduced with prior permission of the Environment Agency.

The views and statements expressed in this report are those of the author alone. The views or statements expressed in this publication do not necessarily represent the views of the Environment Agency and the Environment Agency cannot accept any responsibility for such views or statements.

Author(s):
John Iwan Jones[†], John Davy-Bowker[‡]

Dissemination Status:
Publicly available

Keywords:
Deep Rivers, bioassessment, sampling, methods, airlift, pond net, dredge, RIVPACS, RICT

Research Contractor:
[†]Queen Mary University of London,
Mile End Road,
London, E1 4NS
01929 401892

[‡]Freshwater Biological Association,
The River Laboratory
East Stoke
Wareham
BH20 6BB

Environment Agency's Project Manager:
David Colvill, SEPA

Science at the Environment Agency

Science underpins the work of the Environment Agency. It provides an up-to-date understanding of the world about us and helps us to develop monitoring tools and techniques to manage our environment as efficiently and effectively as possible.

The work of the Environment Agency's Science Group is a key ingredient in the partnership between research, policy and operations that enables the Environment Agency to protect and restore our environment.

The science programme focuses on five main areas of activity:

- **Setting the agenda**, by identifying where strategic science can inform our evidence-based policies, advisory and regulatory roles;
- **Funding science**, by supporting programmes, projects and people in response to long-term strategic needs, medium-term policy priorities and shorter-term operational requirements;
- **Managing science**, by ensuring that our programmes and projects are fit for purpose and executed according to international scientific standards;
- **Carrying out science**, by undertaking research – either by contracting it out to research organisations and consultancies or by doing it ourselves;
- **Delivering information, advice, tools and techniques**, by making appropriate products available to our policy and operations staff.

Steve Killeen

Head of Science

Executive summary

The size and depth of the channel of larger rivers raises important logistic, expense, and safety issues that need to be addressed when developing ecologically sound sampling methods. An approach is required that provides a comprehensive and scientifically defensible evaluation of the condition of water bodies that comprise large rivers. A variety of techniques and strategies are available for collecting a representative sample of the macroinvertebrate community present in river reaches where it is impractical to use standard kick sampling methodology, i.e. reaches where a large proportion (circa. > 40%) of the width is too deep to safely wade. This report reviews the work undertaken to date comparing the performance of available techniques. Any modifications in the design of equipment to be used in routine monitoring from those tested will render the results and recommendations reported here invalid and new tests of performance will be required.

Rivers are variable in space and time: selection of the appropriate sampling strategy and technique will depend on the conditions at the site at the time of sampling, and should not influence measures of ecological quality. Development of a sampling strategy that provides a continuous boundary between deep and shallow sites would provide several advantages.

1. It will not be necessary to develop independent models for deep rivers as they can be integrated into shallow water models.
2. Categorisation of a site as deep or shallow, in terms of the sampling technique to be used, will not influence the ecological status of the site.
3. Deep water reference sites can be grouped along with shallow water reference sites, potentially reducing the number of deep water reference sites required.

Similar issues arise with other contiguous water bodies (lakes, canals, transitional waters). Given the considerable investment and development advances made in the assessment of rivers through RIVPACS (River InVertebrate Prediction and Classification System) and latterly RICT (River Invertebrate Classification Tool), a pragmatic approach would be to develop tools and sampling techniques for contiguous water bodies so that they provide a continuous boundary with RIVPACS.

Across Europe to date there has been a lack of consistency in sampling deep rivers, both within and between member states; the methodologies adopted tend to be selected on a regional or ad hoc basis. A consistent method is required for the UK.

Aspects of the performance and suitability of the available techniques for sampling deep rivers have been rigorously tested under a wide range of environmental conditions encompassing many of the deep river types found in the UK and Republic of Ireland. Unless the design of equipment is modified from that used in these tests, it is recommended that there is no need for further comparative testing of deep river sampling methods.

Grabs do not perform well where substrates are coarse or velocities high. Colonisation samplers using artificial or natural substrata, light traps, and traps for catching drifting invertebrates tend to be inefficient and selective. As none of these techniques are likely to produce samples comparable to those collected using a standard RIVPACS kick sample they are not recommended for classification monitoring in the UK.

The US Environmental Protection Agency has adopted a strategy of sampling the shallow margins (<1 m) of deep rivers. However, samples collected from the margins of deep rivers differ from those collected in the main channel. Wide rivers cannot be effectively sampled at the margin alone as the high scoring mid-channel fauna are overlooked. Such a strategy would also not be compatible with the RIVPACS methodology used to assess shallow rivers. It is also recommended that agencies replace monitoring activities based on sampling accessible areas of deep rivers with methods that provide a sample that represents all habitats, shallow and deep, as soon as is feasibly possible. It is recommended that a strategy for routine monitoring is developed that samples both mid-channel and marginal habitats.

Whilst apparently easy to use and the preferred method amongst many agency staff, objective statistical comparisons indicate that the long-handled pond net under-performs in terms of recovering available BMWP Scoring Taxa from wide deep rivers and should be discounted as a reliable sampling method for deep-water benthos from wider rivers. As a distinct deep water

community appears to occur in wider rivers, use of the long-handled pond-net should be restricted to narrow deep water courses (provisionally < 15 m wide ditches, etc) and an alternative deep-water protocol used in wider channels. Use of the long-handled pond net in wide rivers will result in sites being misclassified as being poorer quality than their true potential.

The light naturalists' dredge has no power to detect differences in quality and, thus, is effectively useless. It is recommended that the light naturalists' dredge is not used for classification monitoring. Heavier dredges (e.g. medium naturalist's dredge) perform better than the light naturalists' dredge, but use by throwing from the bank has health and safety implications: use of heavier dredges by towing from a boat may also be unsafe and would also require full field testing before any conclusions could be drawn.

The airlift provides better representation, is more sensitive and precise, and is more cost-effective (in terms of processing) samples than other methods routinely used for monitoring. It is also recommended that a strategy for reference sample collection and routine monitoring of deep rivers is developed where mid-channel samples are collected with an airlift and combined with sweep samples from the margin collected with a pond net. The design of such an airlift should follow that of the Yorkshire pattern airlift which has been tested.

On the basis of this review the following recommendations for deep water sampling are made:

1. A specific and standardised deep water sampling technique is used in all sites where a large proportion (provisionally > 40%) of the width is too deep to safely wade.
2. Samples should not be collected from shallow patches of deep rivers.
3. Dredges are not recommended for routine monitoring.
4. For narrow deep water courses (provisionally less than 15 m average width subject to verification later in this project) samples should be collected with a long-handled pond net.
5. For wide deep water courses (provisionally greater than 15 m average width), samples from the channel should be collected with a Yorkshire pattern airlift.
6. A sample from the margin (equivalent to the 1 minute search of the shallow water technique) should be combined with an airlift or long-handled pond net sample from the channel.
7. The high precision of Yorkshire pattern airlift samples presents an opportunity to counterbalance the increased costs of sample collection with a reduced sampling frequency for deep rivers.

Acknowledgements

We would like to thank Ben McFarland[1] and John Murray-Bligh of the Environment Agency for their help in developing the research proposal for this work. We would also like to thank Rachel Benstead, Chris Extence, Alice Hiley, Tim Jones, Geoff Phillips and Shelagh Wilson (Environment Agency), David Colvill (Scottish Environment Protection Agency) and Imelda O'Neill (Northern Ireland Environment Agency) for their very useful help and comments at the start up meeting and for comments on the report.

[1] Now with the RSPB (Minsmere).

Contents

1	**Introduction**	**1**
1.1	Deep Rivers and Other Water Bodies	2
1.2	Strategies Used for Sampling Deep Rivers	5
2	**Comparisons of techniques for sampling macro-invertebrates from deep rivers**	**7**
2.1	Elliott, Tullet and Elliott (1993) and works therein	10
2.2	Benjamin (1998)	13
2.3	Wright, Clarke, Gunn, Blackburn and Davy-Bowker (1999)	14
2.4	Bass, Wright, Clarke, Gunn, and Davy-Bowker (2000)	17
2.5	Blocksom and Flotemersch (2005)	27
2.6	Blocksom and Flotemersch (2008)	28
2.7	Neale, Kneebone, Bass, Blackburn, Clarke, Corbin, Davy-Bowker, Gunn, Furse and Jones (2006)	29
3	**Conclusions**	**42**
4	**Recommendations**	**44**
References		**45**
List of abbreviations		**49**

Table 1 Sampling methods for deep river sites employed by the EA, EHS, SEPA and EPA (Republic of Ireland), each by area, and Europe/USA by country. NR indicates not routinely, ± indicates discontinued, and brackets the number of countries using that methodology. (Data for UK agencies from 1999, for other countries from 2005). 6
Table 2 Summary of qualitative samplers suitable for different types of substrata in deep rivers (RIVPACS substrate categories given in brackets). + = sampler is suitable; F = sampler sometimes fails. Airlift samplers used at an airflow >200 L min^{-1}. (Data from Table 4 in Drake and Elliott 1982). 12
Table 3 Responses to questions on some of the practical advantages and disadvantages of alternative procedures for sampling in deep water (Wright *et al*., 1999). Note that the numbers below include non-routine samples. (Figures in brackets indicate responses from non-Agency laboratories). 15
Table 4 Mean and standard deviation (SD) of NTAXA, BMWP Total Score and ASPT, by site for the four techniques tested. Additional replicates for different operators shown for the Airlift and Medium Naturalists' Dredge. 19
Table 5 Comparison of time (hours) and the equivalent number of sample replicates required to recover 80% of the BMWP Scoring Taxa recorded at each site by the deep-water sampling methods tested. (Fastest options highlighted). Note variable results between BAMS series. N/A denotes the yield cannot reach 80% of the recorded taxa. 23
Table 6 Occurrence of taxa confined to deep-water samples (n - number of sample replicates, out of 18, in which the taxon was present). 25
Table 7 Comparison of the NTAXA recorded from deep-water samples, margin pond net samples and combined methods at each site. The combined methods yielding the highest NTAXA are highlighted 26
Table 8 The percentage of total variance attributable to replicate samples and total within site variance of the US EPA large river methodology for shallow (thalweg depth < 4 m) and deep (thalweg depth > 4 m) river sites compared to measures for standard RIVPACS kick samples as tested by Clarke et al. (2006) and the most precise deep river method tested by Neale et al. (2006). Note different metrics are used in the US and UK studies. 28
Table 9 Environmental Conditions at the Sites used in the NS Share Project. 30
Table 10 Estimates of sources of variance in BMWP Score, NTAXA and ASPT for each of the field sampling techniques (airlift, dredge, margin and LHPN). *, ** and *** denote site or operator variance component was statistically significant in ANOVA tests at the 0.05, 0.01 and 0.001 test probability level. 38
Table 11 Comparison of the field sampling techniques (airlift, dredge, margin and LHPN) for sampling processing cost (time in minutes; number of samples shown in brackets) to achieve a sampling variance of less than Q% (20% or 10%) of the total variance amongst all sites in terms of BMWP Score, NTAXA, ASPT, and all 3 metrics. σ_I^2 and σ_W^2 denote between- and within- site variance estimates. 39

Figure 1 Schematic diagrams to illustrate the mode of operation of techniques for collection of samples from shallow, a) standard kick sample, and deep rivers, b) marginal sweep, c) long-handled pond net, d) dredge, e) airlift. 4
Figure 2 Comparison of mean sample sort times between sampler types and sites. 18
Figure 3 Mean BMWP Score for each sampling method and site. 21
Figure 4 Mean ASPT for each sampling method and site. 21
Figure 5 Smoothed taxon accretion curves indicating the predicted NTAXA found in any single, pair, 3, 4, 5, or 6 random samples (out of the total of six replicate samples taken by that method) at each site surveyed using the airlift, medium naturalists' dredge and long-handled pond net. Flattening of curves to a plateau indicates the maximum NTAXA retrievable with that method and the number of replicates required to achieve this. 22
Figure 6 Relationship between macrophyte cover and index scores of 1 minute marginal sweep samples at six deep river sites. 24
Figure 7 Influence of technique on the time taken to sort the samples collected with the four deep water techniques tested. Mean values shown ±SE. Different letters indicate significant differences among mean values as identified by Tukey's test, shared letters indicate no significant difference. 30
Figure 8 Influence of technique and site on a) total BMWP score b) NTAXA and c) ASPT of the samples collected with the four deep water techniques tested. Mean values shown ±SE. Different letters indicate significant differences among mean values as identified by Tukey's test, shared letters indicate no significant difference. 32
Figure 9 Matrix showing correlation between BMWP scores of the four deep water techniques, using pairs of matched replicates from the same site reach. R is shown in the top right hand corner for each combination. 33
Figure 10 Matrix showing correlation between NTAXA of the four deep water techniques, using pairs of matched replicates from the same site reach. R is shown in the top right hand corner for each combination. 34
Figure 11 Matrix showing correlation between ASPT of the four deep water techniques, using pairs of matched replicates from the same site reach. R is shown in the top left hand corner for each combination. 35
Figure 12 Influence of a) the width of the river channel and b) the depth of the centre of the river channel on relative performance of airlift and margin techniques in terms of ASPT. Results of interaction between technique and width and depth from Ancova shown; this interaction indicates differences in relative performance if significant. 36

1 Introduction

As water moves through the landscape so it tends to accumulate in channels which form a merging network of streams and rivers. With distance from source, as headwaters coalesce, rivers tend to become larger both in terms of physical dimensions and discharge. The functional types and distribution of the organisms present changes as the physical dimensions and discharge increase. This increase in size presents challenges when trying to assess biological condition. Although relatively small in size and discharge, headwaters comprise a far greater proportion of the length of rivers. Hence, most sampling programmes have been developed for headwaters where the whole width of the river bed is relatively accessible by wading. In larger rivers the size and depth of the channel raises important logistic, expense, and safety issues that need to be addresses when developing ecologically sound sampling methods.

The difficulties of sampling large (or more precisely, and hereafter, deep) rivers have led to national strategies that have not included deep rivers in assessments, only sampled shallow sections or patches of deep rivers, or have based assessments on easily sampled quality elements. None of these approaches provide a comprehensive and scientifically defensible evaluation of the condition of water bodies that comprise deep rivers. Furthermore, such approaches do not comply with the requirements of the Water Framework Directive. Techniques need to be developed for the assessment of large and deep rivers in the UK.

This report constitutes the first of two which examine the issues concerning the assessment of deep rivers in the UK.

The first report will:

i) Review the results of previous deep-water methods comparison studies, including practical aspects of deep river sampling.
ii) Make recommendations on the preferred deep water sampling method(s) and the threshold between methods for sampling wadeable and deep rivers.
iii) Examine the potential discontinuities in RIVPACS predictive models that might arise from the methods used to collect reference samples.

The second report will:

iv) Identify existing RIVPACS sites that have been inappropriately sampled (given their depth), examine the distribution of deep water reference sites in the current RIVPACS model, and suggest replacement sites.
v) Evaluate the suitability of the classification metrics EQR ASPT and EQR NTAXA for deep rivers.
vi) Examine the potential need for additional environmental variables to adequately discriminate deep rivers in RIVPACS predictive models.
vii) Produce clear guidelines for sampling deep rivers for inclusion in future Environment Agency sampling manuals
viii) Undertake an ergonomic assessment of airlift sampling.
ix) Provide a specification for a ‘standard’ airlift sampling device.
x) Provide a specification and a costed work programme for a Phase II project to collect new deep river samples and build new RIVPACS model(s) based on samples collected using standardised deep water sampling methods.

1.1 Deep Rivers and Other Water Bodies

A variety of techniques and strategies are used for collecting macroinvertebrate samples from deep rivers, although the conditions where they are applied vary according to the definition of a deep river that is adopted. Here we define deep rivers as those river reaches where it is impractical to use standard kick sampling methodology (Murray-Bligh *et al.* 1997) to collect a representative sample of the macroinvertebrate community present in that reach. This is defined as river reaches where a large proportion (circa. > 40%) of the width is too deep to wade safely. Other definitions of large rivers are available based on stream order, discharge or other dimensions, but the practical definition used here addresses the key aspect that differentiates deep rivers from shallow headwaters in terms of the collection of invertebrate samples.

Even this definition presents challenges when differentiating between deep and shallow rivers. Rivers are variable in both space and time: depth can vary markedly between spate and drought conditions. Furthermore, the hydrogeomorphic processes of erosion and deposition alter the morphology of the river bed over time; shallow access points may appear or disappear over time.

Selection of the appropriate sampling strategy and technique must therefore depend on the conditions at the site at the time of sampling.

The variable nature of rivers accentuates a further challenge for the design of sampling strategies and choice of techniques for sampling deep and shallow sites: it is of fundamental importance that the choice of technique does not lead to a different assessment of quality. To obviate the requirement for inter-calibration between deep and shallow sites it is desirable for the samples collected using the recommended deep water technique to be comparable with those collected with the standard shallow water technique, thus providing a continuous transition between deep and shallow rivers. Development of a sampling strategy that provides a continuous boundary would provide several advantages.

1. It will not be necessary to develop independent classification tools for deep rivers as they can be integrated into shallow water models.
2. Classification of a site as deep or shallow, in terms of the sampling technique to be used, will not influence the ecological status of the site.
3. Deep water reference sites can be classified along with shallow water reference sites, potentially reducing the number of deep water reference sites required.

Further issues of inter-calibration arise as a consequence of landscape: deep rivers are not discrete units within the landscape but are interconnected with lakes, canals, and transitional waters, often in a graded fashion where there is no marked boundary between one water body type and the next. Where rivers are impounded, for example by low-head dams, small hydroelectric facilities or navigational dams, the river may retain much of the structure of a flowing river ecosystem yet be lentic, whereas riverine features are lost when large rivers flow into larger lakes and reservoirs. The cut off between these two water body types (i.e. deep rivers and lakes) will need definition. The same is true of transitional waters: deep rivers flow into estuaries and in the absence of artificial structures (weirs, sluices, etc.) the transition is rarely discrete. Engineered channels, canals and canalised rivers are frequent features of the lower end of catchments, often interconnected with deep rivers, again there needs to be a clear discrimination between water body types in the field.

It is of the utmost importance that the choice of technique used at a site does not influence the assessment of quality of different contiguous water bodies: a continuous boundary between contiguous water body types would be preferable, although inter-calibration is an alternative approach.

Given the considerable investment and developmental advances made in the assessment of rivers through RIVPACS (River Invertebrate Prediction and Classification System) and latterly RICT (River InVertebrate Classification Tool), *a pragmatic approach would be to develop tools and sampling techniques for contiguous water bodies so that they provide a continuous boundary with RIVPACS.*

The sampling methodology developed for use at shallow river sites (a 3 minute timed kick/sweep sample with a standard FBA pond net fitted with a 1.5 m handle plus a 1 minute search, sampling all habitats from a site typical of the reach in proportion to their occurrence: see Figure 1) provides a

sample that is representative of the river reach as a whole and is comparatively simple, with the result that a high degree of standardisation is possible (McGarrigle *et al.* 1992; Murray-Bligh *et al*. 1997). In addition, much effort has been devoted in the UK to documenting and reducing sources of error from sampling variation, sorting and identification in order to improve the precision of the technique (Dines and Murray-Bligh, 2000; Clarke *et al.* 2002). In contrast, sampling deep waters is inherently more difficult, hazardous and time-consuming. The biologist has much less control of the sampling device (see Figure 1) and, as a consequence, it is difficult to sample all invertebrate habitats in proportion to their occurrence.

Following the RIVPACS methodology, techniques for deep rivers should as far as possible sample all habitats (marginal and benthic) in proportion to their occurrence. As well as providing samples comparable with those from shallow river sites, this approach would provide a comprehensive assessment of the deep river site that is likely to include the taxa most sensitive to a range of stressors. An inaccurate assessment of quality will be given if parts of the community are missed at a site simply because they are out of range of the technique used.

For the Environment Agency 2000 GQA survey (England & Wales), use of a long-handled pond net (a standard FBA pond net with the 1.5 m long handle modified so that extensions can be fitted to increase the length to 4 m: see Figure 1), the medium naturalists' dredge or the Yorkshire pattern airlift were recommended for deep water sites (See Figure 1. For detailed specification of the design of equipment see Murray-Bligh *et al.* 1997, however, note that in this work the standard pond net with 1.5 m long handle for use in shallow waters is referred to as the standard FBA long-handled pond net). A variety of methods are in regular use for the assessment of deep-water sites across Britain and Ireland, with the methodology adopted determined at a regional level or by the individual collecting the sample (see Table 1).

However, some methods for sampling benthos, such as dredges and airlifts, are more time consuming than the standard pond net technique and some require several people, resulting in increased costs. A protocol on standard sampling effort has yet to be defined for deep river devices. Furthermore, the representation of the benthic community using such devices relative to that achieved using the standard shallow river pond net technique, has not been fully assessed so that the influence of the choice of technique (deep versus shallow technique) on the assessment of a site remains unknown.

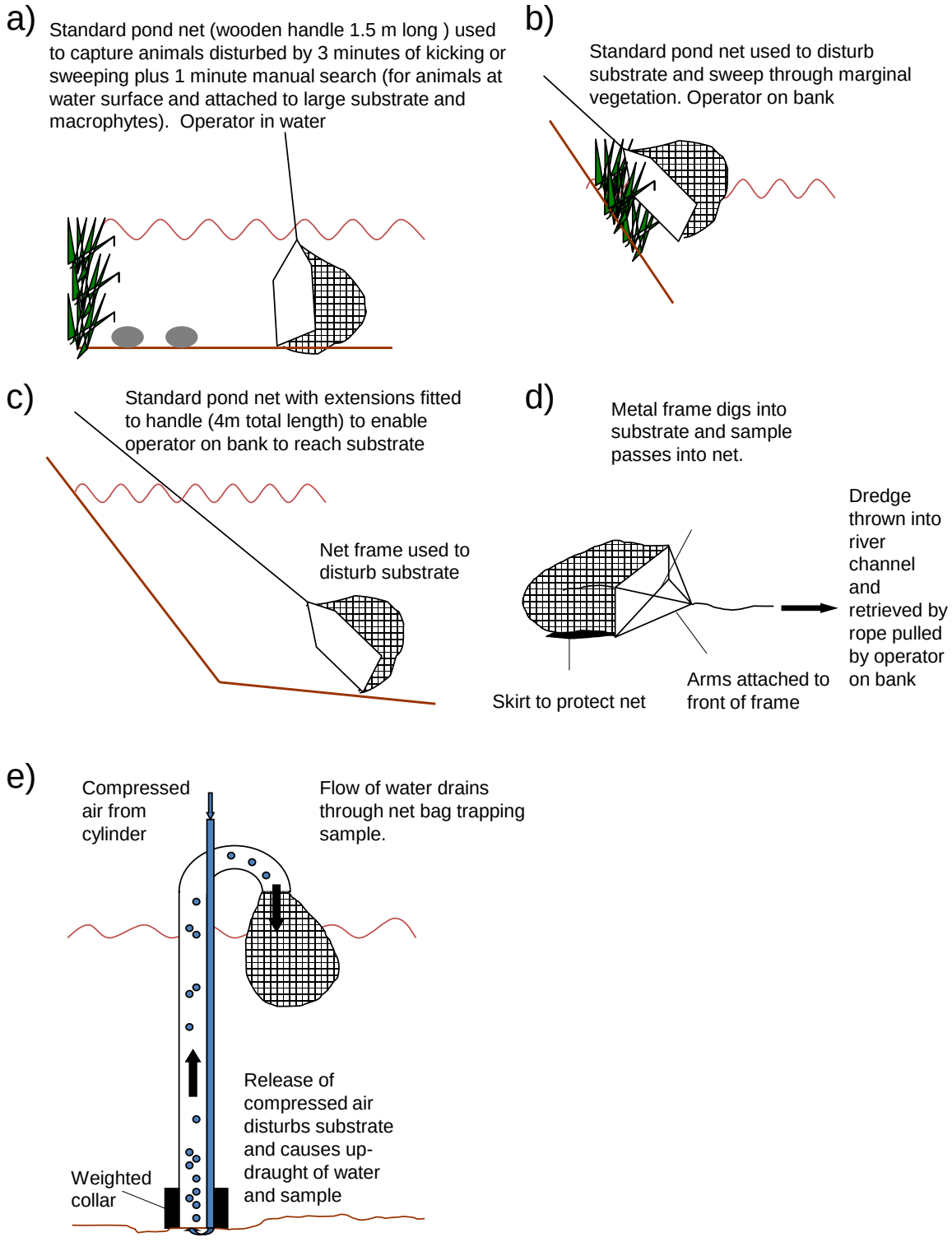


Figure 1. Schematic diagrams to illustrate the mode of operation of techniques for collection of samples from shallow, a) standard kick sample, and deep rivers, b) marginal sweep, c) long-handled pond net, d) dredge, e) airlift.

1.2 Strategies Used for Sampling Deep Rivers

Jones, Bass and Davy-Bowker (2005) collated information from the UK, the Republic of Ireland and mainland Europe on the techniques used to sample deep rivers (Table 1). This information was derived from a questionnaire sent by Wright *et al.* (1999) to EA, SEPA and (at the time) EHS regional offices, and by Jones *et al.* (2005) to workers in the EPA of the Republic of Ireland, and to workers responsible for assessments of water quality in other European member states. Across Europe to date there has been a lack of consistency in sampling deep rivers, both within and between member states; the methodologies adopted tend to be selected on a regional or *ad hoc* basis. The most frequently used techniques are a sweep/pond net or dredge sampler operated from the bank. However, several organisations use techniques that are deployed from boats, namely grabs, airlifts, artificial substrates, freeze coring or scuba diving.

Subsequently, Flotemersch, Stribling, and Paul (2006) produced guidelines for the bioassessment of non-wadeable streams and rivers by the Environmental Protection Agency in the USA. Here, the Large River Bioassessment Protocols, a hybrid of USEPA-EMAP (Lazorchak *et al.*, 2000), USEPA-RBP (Barbour *et al*. 1999) and USGS-NAWQA (Moulton *et al*. 2002) sampling methods, recommend that macroinvertebrates are collected from deep rivers by 6 sweep samples distributed from the edge of water to the mid-point of the river or until depth exceeds 1 m, each 0.5 m in length, using a D frame net (500 µm mesh, width 0.3 m). Each sweep is to cover 0.15 m^2 of substrate; therefore, six sweeps cover an area of 0.9 m^2. The six sweeps are to be allocated in proportion to occurrence of the available habitat within the sample zone (e.g. snags, macrophytes, cobbles). If water at a site is more than 1 m deep at the water's edge, the six sweeps are to be collected from a boat if possible.

The previous Environmental Monitoring and Assessment Program (USEPA-EMAP; Lazorchak *et al.* 2000) recommendations for non-wadeable rivers, were to collect kick net samples from shallow areas (< 1m) near the bank of the river using an oblong net 50 x 30 cm. Two kick samples were to be collected from each of eleven transects. In addition two daytime drift net samples were to be collected from the downstream end of the defined reach, positioned and retrieved by boat. Artificial samplers (Hestor-Dendy, rock-filled baskets) had been recommended in the past and were used for routine monitoring in some states (e.g. Ohio EPA).

Table 1. Sampling methods for deep river sites employed by the EA, EHS, SEPA and EPA (Republic of Ireland), each by area, and Europe/USA by country. NR indicates not routinely, ± indicates discontinued, and brackets the number of countries using that methodology. (Data for UK agencies from 1999, for other countries from 2005).

Region	Area	Sweep	Distur-bance	Dredge	Airlift	Grab	Marginal Kick	Search	Artificial Substrate	Other
Anglian	Eastern	+	+	+						
	Central	+	+							
	Northern	+	+			NR				
Midland	Upper Severn	+								
	Lower Severn	+	+	+						
	Upper Trent	+	+	+			+			
	Lower Trent	+	+							
North East	Dales	+		+	+					
	Ridings	+	+	+	+					
	Northumbria	+			NR					
North West	Northern	+	+							
	Central	+	+							
	Southern									
Southern	Kent	+	+	+						
	Sussex	+	+	+						
	Hants & IOW	+	+	+						
South West	Cornwall	+		+						
	Devon	+		+						
	North Wessex	+	+	+					NR	
	South Wessex	+		+						
Thames	North East	+	+	+				+		
	South East	+	+							
	West	+	+	+	+					
Welsh	North	+	+	NR						
	South East	+	+							
	South West	+		NR						
SEPA	North	+				+				
	Dumfries			+		+				
	East Kilbride	+	+							
EA N. Ireland		+	+	NR			+			
EPA	Southern	+	+	±		±				
	Dublin	+	+							
Austria		+		+	+	+			+	Freeze core Scuba
Czech Republic		+		+	+					
France				+		+			+	
Germany		+					+			
Greece		+	+			+				Sweep + grab
Latvia						+				
USA		+					±	+	±	Drift
Totals		35 (9)	23 (5)	22 (7)	6 (3)	8 (7)	4 (4)	2 (2)	4 (4)	

2 Comparisons of techniques for sampling macro-invertebrates from deep rivers

Several works have reviewed the use and efficiency of various techniques for sampling macroinvertebrates from deep waters. These works have varied in their approach from descriptive, qualitative comparisons to quantitative, statistical comparisons. The more comprehensive works have compared the techniques across a wide range of conditions, encompassing most of the physical conditions that are likely to be encountered when sampling deep rivers.

It should be stressed most strongly that any information produced in these comparisons, and therefore the recommendations that arise from them, are specific to the exact design specifications and implementation of the equipment used. Any deviation from the design of the equipment used in these works will negate the results provided. Should a modified design or implementation strategy be used for any of the techniques reviewed here, new assessments of their performance will be required before they can be used for routine monitoring. All design details should be held constant as far as possible, extra care should be taken to maintain the weight and dimensions of the operational parts of samplers (jaws, pipes etc.) including the size and mesh size of nets.

It is strongly recommended that the design specifications and implementation of the sampling equipment to be used within this project (ergonomic testing, new reference sample collection), and for routine monitoring of deep rivers thereafter, should be consistent with that used previously. Otherwise, the result of previous works will not be relevant and new assessments of performance will be required.

As there is a current lack of standardised protocols for sampling macroinvertebrates in UK deep waters, here we review the merits of the various deep-water sampling devices available. Several of the techniques reviewed here are not compatible with the RIVPACS approach: they are included here for completeness, particularly as some authors have advocated their use for sampling deep rivers and in some cases are used for bio-assessment of deep rivers in other countries.

Works considered to be of particular relevance to deep-water sampling are:

Bass, J. A. B., Wright, J. F., Clarke, R. T., Gunn, R. J. M. & Davy-Bowker, J. (2000) Assessment of sampling methods for macroinvertebrates (RIVPACS) in deep watercourses. Environment Agency R&D Technical Report E134. 57pp.

Benjamin, J. (1998) A comparative study of methods for sampling macroinvertebrates in Sussex Rifes. Unpublished report to Environment Agency, Southern Region. 103pp.

Blocksom, K.E. and Flotemersch, J.E. (2005) Comparison of macroinvertebrate sampling methods for nonwadeable streams. *Environmental Monitoring and Assessment*, **102**, 243–262.

Blocksom, K.A. and Flotemersch, J.E. (2008) Field and laboratory performance characteristics of a new protocol for sampling riverine macroinvertebrate assemblages. *River Research and Applications*, **24**, 373-387.

Bretschko, G. and B. Schönbauer (1998) Quantitative sampling of the benthic fauna in a large, fast flowing river (Austrian Danube). *Archiv für Hydrobiologie* Supplement, **115**, 195-211.

Collier, K.J., Hamer, M. and Chadderton, W.L. (2009) A new substrate for sampling deep river macroinvertebrates. *New Zealand Natural Sciences,* 34, 49-61.

Czerniawska-Kusza, I. (2004) Use of artificial substrates for sampling benthic macroinvertebrates in the assessment of water quality of large lowland rivers. *Polish Journal of Environmental Studies,* **13**, 579-584.

Depauw, N., Lambert, V., Vankenhove, A. and Devaate, A. B. (1994) Performance of 2 Artificial Substrate Samplers for Macroinvertebrates in Biological Monitoring of Large and Deep Rivers and Canals in Belgium and The Netherlands. *Environmental Monitoring and Assessment,* **30**, 25-47.

Downing, J. A. and Rigler, F. H. (eds.) (1984) *A manual on methods for the assessment of secondary productivity in fresh waters.* IBP Handbook No. **17**. Blackwell, Oxford.

Drake, C. M. and Elliott, J. M. (1982) A comparative study of three air-lift samplers used for sampling benthic macro-invertebrates in rivers. *Freshwater Biology*, **12**, 511-533.

Drake, C. M. and Elliott, J. M. (1983) A new quantitative air-lift sampler for collecting macroinvertebrates on stony bottoms in deep rivers. *Freshwater Biology*, **13**, 545-559.

Elliott, J. M. and Drake, C. M. (1981a) A comparative study of seven grabs used for sampling benthic macroinvertebrates in rivers. *Freshwater Biology*, **11**, 99-120.

Elliott, J. M. and Drake, C. M. (1981b). A comparative study of four dredges used for sampling benthic macroinvertebrates in rivers. *Freshwater Biology*, **11**, 245-261.

Elliott, J. M., Drake, C. M. and Tullett, P. A. (1980). The choice of a suitable sampler for benthic macroinvertebrates in deep rivers. *Pollut. Rep. Dep. Environ. U.K*. No. **8**, 36-44.

Flannagan, J. F. (1970) Efficiencies of various grabs and corers in sampling freshwater benthos. *Journal of the Fisheries Research Board of Canada*. **27**, 1691-1700.

Flotemersch, J. E., Blocksom, K. A., Hutchens, J. J. and Autrey, B. C. (2006) Development of a standardized large river bioassessment protocol (LR-BP) for macroinvertebrate assemblages. *River Research and Applications* **22,** 775-790.

Flotemersch, J. E., Stribling, J. B. and Paul, M. J. (2006) *Concepts and approaches for the bioassessment of non-wadeable streams and rivers*. EPA 600-R-06-127. US Environmental Protection Agency, Cincinnati, Ohio.

Flotemersch, J. E., Blocksom, K. A., Hutchens, J. J. and Autrey, B. C. (2004) *Association among invertebrates and habitat indicators for large rivers in the Midwest: how sampling distance, point-sampling of habitat, and subsample size effect measures of large river macroinvertebrate assemblages*. EPA-600-R-04-177. US Environmental Protection Agency, Cincinnati, Ohio.

Flotemersch, J. E., Autrey, B. C., and Cormier, S. M. (2001) *Comparisons of boating and wading methods used to assess the status of flowing waters.* EPA 600-R-00-108. US Environmental Protection Agency, Cincinnati, Ohio.

Haase, P., Lohse, S., Pauls, S., Schindehuette, K., Sundermann, A., Rolauffs, P. and Hering D. (2004) Assessing streams in Germany with benthic invertebrates: development of a practical standardised protocol for macroinvertebrate sampling and sorting. *Limnologica,* **34**, 349-365.

HMSO (1984) *Methods of biological sampling: Sampling of benthic macroinvertebrates in deep rivers 1983. Methods for the examination of waters and associated materials.* HMSO, London. 16pp.

Herrig, H. (1975) Der Bodensauger – ein neuartiges Gerät zur Entnahme von Sohlenproben aus großen Fließgewässern. *Dt. Gewässerkdl. Mitt*., **19**, 104-107.

Humpesch, U. H and Elliott, J. M. (eds.) (1990) Methods of biological sampling in a large, deep river - the Danube in Austria. *Wasser Abwasser (Suppl.)* **2/90**, 83pp.

Humpesch, U. H. and Niederreiter, R. (1993) Freeze-core method for sampling the vertical-distribution of the macrozoobenthos in the main channel of a large deep river, the river Danube at river kilometer 1889. *Archiv für Hydrobiologie* Supplement, **101,** 87-90.

Humphries, P., Growns, J. E., Serafini, L. G., Hawking, J. H., Chick, A. J and Lake, P. S (1998) Macroinvertebrate sampling methods for lowland Australian rivers. *Hydrobiologia* **364** (2), 209-218.

Jackson, M. J. (1997) *Sampling methods for studying macroinvertebrates in the littoral vegetation of shallow lakes*. Broads Authority, Norwich.

Jones, J.I., Bass, J.A.B. and Davy-Bowker, J. (2005) *Review of methods for sampling invertebrates in deep rivers*. North South Shared Aquatic Resource (NS Share) Interim Report. 46pp

Lazorchak, J.M., Hill, B.H., Averill, D.K., Peck, D.V. and Klemm D.J. (2000) *Environmental Monitoring and Assessment Program -Surface Waters: Field Operations and Methods for Measuring the*

Ecological Condition of Non-Wadeable Rivers and Streams. U.S. Environmental Protection Agency, Cincinnati OH

Mackey, A. P., Cooling, D. A. and Berrie, A. D. (1984) An evaluation of sampling strategies for qualitative surveys of macro-invertebrates in rivers, using pond nets. *Journal of Applied Ecology*, **21**, 515-534.

McGarrigle, M. L., Lucey, J. and Clabby, K. J. (1992) Biological assessment of river water quality in Ireland. In: *River Water Quality – Ecological Assessment and Control.* 371-393, Commission of the European Communities, EUR 14606 EN-FR, Luxembourg, 751pp.

Murray-Bligh, J. A. D., Furse, M. T., Jones, F. H., Gunn, R. J. M., Dines, R. A. and Wright, J. F. (1997) *Procedure for collecting and analysing macroinvertebrate samples for RIVPACS.* Institute of Freshwater Ecology & Environment Agency, 155pp.

Neale, M.W., Kneebone, N.T. Bass, J.A.B., Blackburn, J.H., Clarke, R.T., Corbin, T.A., Davy-Bowker, J., Gunn, R.J.M., Furse, M.T. and Jones J.I. (2006) *Assessment of the Effectiveness and Suitability of Available Techniques for Sampling Invertebrates in Deep Rivers*. North South Shared Aquatic Resource (NS Share) Final Report T1(A5.8) – 1.1. 97pp.

Ofenböck, G. and Moog, O. (2000) The Danube-Net-Basket-Sampler - a simple but effective sampling gear for sampling benthic invertebrates in deep and large stony rivers. *Archiv für Hydrobiologie* Supplement, **115**, 557-573.

Pearson, R. G., Litterick, M. R. and Jones, N. V (1973) An air-lift for quantitative sampling of the benthos. *Freshwater Biology*, **3**, 309-315.

Pehofer, H. E. (1998) A new quantitative air-lift sampler for collecting invertebrates designed for operation in deep, fast-flowing gravelbed rivers. *Archiv für Hydrobiologie* Supplement, **101**, 213-232.

Petermeier, A. & Schöll, F. (1996) *Das hyporheische Interstitial der Elbe – Methodenrecherche*. Bundesanstalt für Gewässerkunde, Koblenz. BfG-1038.

Swift, M. C., Canfield T. J. and LaPoint, T. W. (1996) Sampling benthic communities for sediment toxicity assessments using grab samplers and artificial substrates. *Journal of Great Lakes Research* **22**, 557-564.

Turner, A. M. and Trexler, J. C. (1997) Sampling aquatic invertebrates from marshes: evaluating the options. *Journal of the North American Benthological Society,* **16**, 694-709.

Voshell, J. R., Hiner, S. W. and Layton, R. J. (1992) Evaluation of a benthic macroinvertebrate sampler for rock outcrops in rivers. *Journal of Freshwater Ecology* **7**, 1-6.

Wagner, F., Zimmermann-Timm, H. and Schonborn, W. (2003) The Bottom Sampler - a new technique for sampling bed sediments in streams and lakes. *Hydrobiologia* **505**, 73-76.

Williams, P., Biggs, J., Whitfield, M., Corfield, A., Fox, G. and Adare, K. (1998) *Biological techniques of still water quality assessment. 2. Method development.* Report to the Environment Agency, R&D Technical Report E56. 158pp.

Wright, J. F., Clarke, R. T., Gunn, R. J. M., Blackburn, J. H. and Davy-Bowker, J. (1999) Testing and further development of RIVPACS – Phase 3. Development of new RIVPACS methodologies . Stage 1. 138pp. Environment Agency.

Wright, J.F., Winder, J.M., Gunn, R.J.M., Blackburn, J.H., Symes, K.L. and Clarke, R.T. (2000) Minor local effects of a River Thames power station on the macroinvertebrate fauna. *Regulated Rivers: Research and Management*. **16**, 159-174.

Although there is apparently a wide array of techniques, many of the different designs are refinements to improve effectiveness when collecting quantitative samples, refinements that are unnecessary for routine bio-assessment. The RIVPACS method for assessment of shallow rivers relies on species occurrence or semi-quantitative samples of macroinvertebrates, with effort put into sampling habitats in proportion to their occurrence rather than to produce quantitatively accurate samples from each habitat. Methodologies for routine bio-assessment of deep rivers should be consistent with the RIVPACS methods used in shallow rivers in this respect.

The methodologies available for collecting macroinvertebrate samples from deep waters can be broadly categorized into:

- Nets
- Grabs
- Dredges
- Airlifts
- Freeze-corers
- Artificial substrates
- Light traps

Some of these and allied methodologies have been reviewed for their use in collecting invertebrate samples from amongst macrophytes in lake littorals by Jackson (1997) who concluded from a desk study of their practical use and efficiency that a pond net used from a boat was as good as any other technique, in terms of cost, ease and speed of use, and perceived efficiency of capture of species. Field comparisons of various methods in heavily vegetated sites (Turner & Trexler 1997) also indicated that the sweep net was efficient at describing the community, together with the stovepipe (a cylinder used to isolate a vertical water column combined with the use of sweep netting inside) and a funnel trap. Hester-Dendy artificial substrates, minnow trap, a benthic corer, and a plankton net were ineffective. It was noted that the number of species recorded was proportional to the number of individuals, with the techniques most effective at capturing individuals producing the best description of the community. Although macrophytes are frequently present in deep rivers, samplers specifically designed for collecting macroinvertebrates from amongst macrophytes are not considered here. Freeze-corers are not considered either, as they are designed to collect quantitative samples of the hyporheos (fauna interstitial within the sediment) and are regarded as unnecessarily complex for regular biomonitoring.

Although artificial substrates are used regularly in some countries (e.g. Austria, France) they are considered selective and require two site visits to collect a sample. Light traps have the same drawbacks, and do not appear to have widespread use. Furthermore, artificial substrates and light traps are not compatible with the RIVPACS. Hence, we will restrict our consideration to nets, grabs, dredges and airlifts.

2.1 Elliott, Tullet and Elliott (1993) and works therein

Elliott, Tullet and Elliott (1993) provide a comprehensive bibliography of designs and comparisons of devices used for sampling benthic invertebrates from the natural substrata of rivers and streams, published by the Freshwater Biological Association, Occasional Publication No. 30 '*A new bibliography of samplers for freshwater benthic invertebrates'*. In the works referred to therein, comparisons were made with respect to quantitative measures of community composition, rather than biomonitoring *per se*: differences in indices and quality assessments were not considered. Current bio-assessment methods for shallow rivers rely on species occurrence or semi-quantitative samples of macroinvertebrates, with more effort put into sampling habitats in proportion to their occurrence than to produce quantitatively accurate samples from each habitat.

The bibliography of Elliott *et al*. (1993) does not include references to colonisation samplers using artificial or natural substrata, or to light traps, or to traps for catching drifting invertebrates, upstream-moving invertebrates and the emerging imagines of aquatic insects. These methods are dealt with by Elliott, Drake and Tullett (1980). As these techniques are not compatible with RIVPACS, the review here will not consider them further.

Therein, the most relevant works are those of Åarefjord (1972), Drake and Elliott (1982), Mackey (1972), Norris (1980), and Pearson, Litterick and Jones (1973), who compared airlift samplers with other techniques. In each of these comparisons the airlifts (of various design) performed well, being comparable to or better than the other techniques tested (dredges, grabs, surbers) in terms of the composition of the fauna collected.

A large number of publications deal with comparisons between grab samplers of various design, namely Wasmund (1932), Kajak (1963), Beeton, Carr and Hiltunen (1965), Gaufin, Harris and Walter (1965), Stańiczykowska (1966), Brinkhurst, Chua and Batoosingh (1969), Prejs (1969), Sly (1969), Hudson (1970), Howmiller (1971), Burton and Flannagan (1973), Milbrink and Wiederholm (1973),

Weber (1973), Holopainen and Sarvala (1975), Thayer, Williams, Price and Colby (1975), Karlsson, Bohlin and Stetson (1976), Bakanov (1979), Baker, Kimball and Bedinger (1977), Andre, Legendre and Harper (1981), and Drake and Elliott (1983). These publications variously state that grab samplers only work effectively in soft, fine grained sediments. Slack, Ferreira and Averett (1986) compared the ponar grab sampler with artificial substrates, and concluded that both were selective in the community that was sampled

At the beginning of the 1980's a comprehensive assessment of seven grabs (Elliott & Drake 1981a), four dredges (Elliott & Drake 1981b) and three airlift samplers (Drake & Elliott 1982) was undertaken by members of FBA staff at the Windermere Laboratory. All equipment tested was suitable for use from a small boat, larger equipment requiring a winch was not tested. This was a prelude to the development of the FBA Airlift sampler (Drake & Elliott 1983), which was capable of taking quantitative samples on substrata ranging from fine gravel (modal size 0.5-4 mm; RIVPACS = sand and gravel) to large stones (modal size 128-256 mm; RIVPACS = cobbles), although it was not recommended for use on mud (RIVPACS = silt & clay).

According to Elliot & Drake (1981a), grabs do not perform well where the substrate is coarse, particularly at sites where the water is deep (more than 1m) and the current is fast (more than 0.5 m s^{-1}). Furthermore, grabs often leak around the moving parts, resulting in loss of the fine fraction during lifting, and this problem is exacerbated by stones or other debris becoming trapped in the jaws and preventing them from closing properly. These problems restrict grabs to soft sediments in sluggish rivers, and exclude them from use in regular biomonitoring where samples must be collected from a range of conditions. Working under difficult conditions in field trials on the Danube (both deep and of high velocity), the Petersen grab and slurp gun (Herrig 1975) consistently performed badly, underestimating many taxa when compared to the FBA airlift and a deep water freeze corer designed for sampling hyporheos from coarse gravelly sediments (Bretschko & Schönbauer, 1998). The airlift consistently produced the most individuals and most species, and was the preferred method. A similar result was found on the River Elbe (Petermeier & Schöll 1996). *Both these European studies recommend the Airlift for routine sampling of deep rivers.*

Using the modified FBA airlift, Pehofer (1998) found significant differences in community composition between deep water samples and samples collected from an adjacent gravel bar using a Hess sampler. This suggests that *samples from only the shallow or only the deep sections of rivers fail to represent community composition as a whole.*

Various dredges were compared by Fast (1968), Elliott and Drake (1981b) and Probert (1984), but largely not with other techniques. Drake & Elliott (1982) included a summary of qualitative and quantitative samplers suitable for different types of substratum in deep rivers. The section of the table dealing with qualitative samplers is reproduced here as Table 2. Note that the original medium naturalist's dredge referred to in Elliott and Drake (1981b) weighed 9 kg. Although a variety of lower weights ranging from 3-7 kg were previously used within the Environment Agency, the 1.5 kg light naturalists' dredge is the model which is currently recommended on health and safety grounds (Rayson 2000); the effectiveness and suitability of this light naturalists' dredge was assessed by Neale *et al*. (2006) – see section 2.7. The Yorkshire airlift, as described in Murray-Bligh *et al.* (1997), is essentially based on the Mackey airlift (Mackey 1972). Hence, table 2 offers a comparison of the two genuine deep water sampling devices in frequent use in the UK, namely dredges and airlifts. In addition, the mini Van-Veen grab, the Ekman grab, and the FBA airlift (not featured in Table 1) are used on occasions throughout Europe. However, each of these last three devices take small, and in the case of the Ekman grab and FBA airlift, quantitative, samples of substratum and, hence, are inappropriate for regular biomonitoring.

The data of Drake and Elliott (1982) summarised in Table 2 indicates that the medium naturalist's dredge is suitable for sampling substrata ranging from gravel to large stones (cobbles). However, it is unsuitable for sampling mud (silt & clay) and sometimes fails when used on river beds with very large stones (boulders). In contrast, the Mackey airlift was suitable for use on a range of substrata ranging from mud (silt & clay) to small stones (pebbles). Hence, these two sampling devices, although individually deficient on mud (silt & clay), medium naturalist's dredge, and large/very large stones (boulders), Mackey airlift, offer overlapping procedures to ensure that the full range of substrata in deep rivers are amenable to qualitative sampling. These genuine deep-water sampling devices appeared to offer the best options for field trials to determine future sampling protocols.

Table 2 Summary of qualitative samplers suitable for different types of substrata in deep rivers (RIVPACS substrate categories given in brackets). + = sampler is suitable; F = sampler sometimes fails. Airlift samplers used at an airflow >200 L min^{-1}. (Data from Table 4 in Drake and Elliott 1982).

	Substratum					
	Mud (Silt & Clay)	Fine Gravel (Sand/ Gravel)	Fine gravel & small stones (Sand/ Gravel/ Pebbles)	Small stones (Pebbles)	Large stones (Cobbles)	Very large stones (Boulders)
Modal particle size (mm)	<0.1	0.5-4	0.5-4 & 16-32	16-32	64-128	128-256
Van Veen grab	+	+	+F			
Ponar grab*	+	+	+F			
Weighted Ponar grab*	+	+	+			
Birge-Ekman grab (pole-operated)	+	+F				
Allan Grab (pole-operated)	+					
Large Naturalist's dredge		+	+	+	+	+F
Medium Naturalist's dredge		+	+	+	+	+F
Irish dredge†		+	+	+	+	+F
Fast dredge†				+	+	+F
Mackey Airlift	+	+	+	+		
Pearson *et al.* Airlift	+	+	+	+		

* Note that the specific design and construction can influence the effectiveness of grabs, with these factors largely dependent upon the manufacturer.

† Note that large numbers of samples must be taken when using the Irish and Fast dredges.

2.2 Benjamin (1998)

Benjamin (1998) compared standardised methods in use within the Environment Agency for sampling the macroinvertebrate fauna of Sussex Rifes (deep drainage ditches) to determine whether the methodology influenced the results and therefore the perceived water quality. Seven techniques involving the use of pond-nets, dredges, grabs and artificial substrates were used at two sites (3-7 m in width). The techniques which collected the widest range of taxa combined with high abundance for a given sampling effort were kick-sweep pond-net, bank sweep pond-net and dredge. In general these methods also produced the highest biotic scores. Nevertheless, there were sometimes substantial differences in the results obtained by these three methods. Overall, the results justified the use of a bank-sweep plus dredge sample because there were large faunal differences between these components and, therefore, both components were required in order to ensure a representative sample of the whole water course. The long-handled pond-net has subsequently been found to be unwieldy to use from a boat (Bass *et al*. 2000).

This suggests that *the retention of a modified shallow-water protocol based solely on a long-handled pond-net should be restricted to very narrow drainage ditches (<15 m mean width) and that an alternative deep-water protocol must be used in wider channels.*

2.3 Wright, Clarke, Gunn, Blackburn and Davy-Bowker (1999)

As well as analysing data from field trials, Wright *et al.* (1999) compared the relative merits of different sampling techniques as perceived by workers involved in regular monitoring by means of a questionnaire. At that time eight techniques were in use within the EA, SEPA and IRTU (marginal sweep, long-handled pond-net, dredge, airlift, modified Van Veen or Ekman grab, marginal kick sample, deep water kick sample and artificial substrates), and these were compared in terms of their perceived ease of use, efficiency, and the time in the field and the laboratory to process the sample. Each technique was given a "score" on a 3 point scale, with workers asked to comment on the advantages and disadvantages of the methods used in their area. The responses are reproduced in Table 3.

A number of clear patterns emerged in the answers to this question on the practical experience of biologists in sampling deep waters. However, it is important to bear in mind that all responses must be viewed in context. Thus, opinions expressed on the ease of use or efficiency of a procedure are based on perception and limited to the context for sampling (i.e. use of a marginal sweep or a dredge can only be appraised in relation to the marginal areas or river bottom respectively).

It should be noted that the workers were only asked to comment on the techniques that they had practical experience of using. For the EA, SEPA and IRTU workers surveyed, the field sampling protocol for use in shallow streams and rivers has been set out in detail (a 3 minute pond-net sample plus one minute manual search; Murray-Bligh *et al.* 1997) and has been shown to offer a reliable basis for comparing the fauna observed at a site with the expected fauna, as determined by a site-specific RIVPACS prediction (Furse *et al.* 1995). In deep watercourses where kick-sampling is inappropriate, the currently applied EA sampling manual, Murray-Bligh *et al.* (1997), recommended the use of a pond-net (with an extension to the handle if necessary) to obtain a sweep sample of the marginal vegetation plus a sample of the fauna from the river bed in the main channel. Details are given for collecting samples from the channel using a dredge or a Yorkshire pattern airlift, coupled with a sample collected from the margin (equivalent to the 1 minute search of a standard kick sample; Murray-Bligh *et al.* 1997). However, the manual indicates that these latter procedures are interim pending full testing and less preferred, due to a perception of the equipment being more difficult to control and possibly less efficient on very soft river beds. Hence, for some techniques the response was limited: it would be unwise to attempt to draw any firm conclusions for any technique where the number of responses was limited.

It should be noted that the perceived performance of the various techniques may not match with their performance when assessed with objective statistical tests.

In general, the marginal sweep technique was perceived as a simple and efficient means of obtaining a BMWP family list for a site that entailed a short time in the field and only moderate time for subsequent laboratory processing.

The long-handled pond-net technique for sampling the river bottom was also regarded as simple to use, but frequently of only moderate efficiency, sometimes involving more time in the field than marginal sweep sampling and moderate time in the laboratory for sample processing.

Dredges were regarded as moderately easy to use in the field and reasonably efficient at collecting the fauna, albeit with a wide range of responses from good, through moderate to poor (NB at the time the light naturalists' dredge was not the recommended technique). Time in field also varied considerably, with a relatively even response from short, through moderate to long. Laboratory processing of dredge samples was more widely regarded as taking a long time. Although the number of responses for airlifts was low, the available information tended to follow a similar pattern to the dredge, with moderate ease of use, efficiency and time in field, followed by long period for laboratory processing of samples. Although additional protocols were listed, the number of responses was very limited.

Table 3. Responses to questions on some of the practical advantages and disadvantages of alternative procedures for sampling in deep water (Wright et al., 1999). Note that the numbers below include non-routine samples. (Figures in brackets indicate responses from non-Agency laboratories).

Sampling Method	Ease of Use	Efficiency	Time in field	Time in lab
Marginal sweep only	simple 14 + (5) moderate 5 complex 1	good 11 + (1) moderate 6 + (4) poor 2	short 12 + (3) moderate 5 + (2) long 1	short 7 moderate 8 + (3) long 4
Disturbance of substrate	simple 7 + (4) Moderate 7 complex 1	good 3 + (2) Moderate 11 + (2) poor 1	short 6 + (3) Moderate 7 + (1) long 2	short 2 Moderate 10 + (3) long 3
Dredge	simple 4 + (1) moderate 8 + (1) complex 6	good 4 moderate 8 poor 6 + (2)	short 7 moderate 6 + (1) long 5 + (1)	short 1 moderate 5 + (1) long 12
Airlift	moderate 2 complex 1	good 1 moderate 1 poor 1	moderate 2 long 1	long 3
Grab	simple (1) moderate (2)	good (1) moderate (1) poor (1)	moderate (1) long (2)	moderate (1) long (1)
Marginal kick	simple 2	good 1 moderate 1	short 1 moderate 1	short 1 moderate 1
Deep water kick	simple 1	good 1	short	short
Artificial substrate	complex 1	poor 1	long 1	moderate 1
Hand search of boulders etc	simple 1	good 1	short 1	short 1

From the answers to the questionnaire, it was apparent that, whereas a majority of Environment Agency biologists had used long-handled pond-nets to sample the river bed in deep rivers, almost as many had used dredges. In contrast, few employed airlifts. The results of the questionnaire also revealed considerable variation in the detailed specification and use of the various devices, providing further evidence that current procedures for deep water sites are poorly standardised.

Wright *et al.* (1999) also detailed a field comparison undertaken at two sites in Yorkshire (Rivers Aire and Calder), comparing marginal sampling and sampling of the benthos by long-handled pond-net, Yorkshire pattern airlift and medium naturalist's dredge. There were some problems with the way the samples were collected: at one site the airlift was deployed from a bridge across the river, and the substratum collected by the dredge was oily ooze (anoxic silt) and contrasted with the stony substratum sampled by the airlift next to the bridge. The area of river-bed sampled by the airlift was somewhat greater than the 5 m trawl taken with the dredge. Hence, drawing firm conclusions from this work is not possible. Nevertheless, at both sites the maximum number of taxa was obtained by combining the results of the margin with those of the airlift sample.

2.4 Bass, Wright, Clarke, Gunn, and Davy-Bowker (2000)

In a series of field trials at six sites across England and Wales, Bass *et al*. (2000) compared the long-handled pond-net, Medium Naturalist's dredge, Mackey/Yorkshire pattern airlift, and marginal sweep. To ensure a broad scope for comparisons between sampling methods, a range of representative deep-water sites across four Environment Agency regions known to support diverse macroinvertebrate communities were selected for the trials (*North East:* Yorkshire Derwent, Yorkshire Ouse. *Anglian:* Great Ouse/New Bedford River, South Drove Drain. *South West:* River Huntspill. *Midlands:* River Severn). A boat was used at all sites, providing a stable platform from which to take the airlift and long-handled pond net samples. The operators who collected the primary samples were experienced in the use of an long-handled pond-net (IFE), medium naturalist's dredge (IFE) and Mackey/Yorkshire pattern airlift (EA). In order to compare the selected methods in a systematic way, the sampling effort and range of habitat types sampled was consistent between each replicate sample. Six replicate samples per technique were collected at each site to provide a robust indication of sample variability, taxon accretion and for comparison of methods.

The prime objective of the study was to compare the performance and yield of the specified deep-water sampling devices. Samples were collected in the same region of riverbed, in an upstream sequence to prevent dislodgement of the fauna and downstream drift into as yet un-sampled river bed and to avoid sampling the same area more than once. Each deep-water replicate sample was restricted to an area of about 1.5 m^2 so that comparable areas of riverbed were sampled by each method. Samples were preserved and processed in the laboratory. Macroinvertebrates were identified to BMWP family level and the abundance of each BMWP family in the replicate sample was counted to maximise accuracy of between-method assessments.

The authors stress that comparisons between methods need to be unambiguous and objective.

A preliminary measurement of inter-operator variability was made for the medium dredge and airlift at one site each using EA staff from local offices.

The field trial included a programme for sampling the watercourse margins with a pond net. Margin sampling and its contribution to site quality assessments required the collection and analysis of separate data series to facilitate interpretation. A further consideration was the comparison of the fauna from deep-water habitats with the fauna in margin habitats.

The field trial examined the potential benefits of:

- a 3-minute pond net sample from the watercourse margins in preference to a 1-minute marginal sample
- sampling the margin zone of one or both banks
- utilising results from both the watercourse margins and mid-channel habitats.

2.4.1 Sampling activity

Deployment and recovery of the boat, carrying sampling equipment and samples took about two hours at each site, with the rest of the day taken up with the extensive sampling activities. On this basis, the more limited sampling activities during routine monitoring will permit sampling to be completed at two or possibly three deep-water sites in a standard working day. This assumes <1 hour travelling time between sites.

2.4.2 Comparison of sample processing time

Two separate steps were involved in sample processing: (1) macroinvertebrate detection and recovery (referred to as sort time) and (2) identification and counting. The sort time for the different

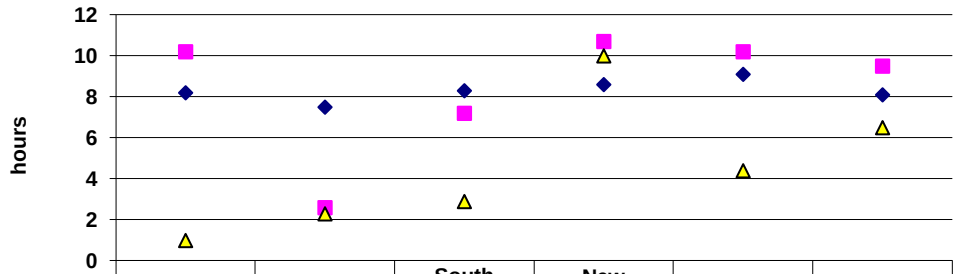
sampling devices and different sites was already identified as an important practical consideration in the assessment of, and subsequent recommendation of a deep river sampling method.

The time taken to sort the macroinvertebrates sampled by each method to BMWP Family absolute abundance was examined for inter-operator variability (Figure 2). It should be emphasised that sample size varied greatly between methods and sites, despite the attempt to obtain each replicate from a consistent area. Mean sort time was around 7 hours per replicate, with overall sort time ranging from 0.3-20 hours.

The time required to recover macroinvertebrates from the deep-water samples was strongly influenced by sample debris volume (reflecting site conditions), the area sampled (as far as possible kept constant) and the characteristics of each sampling method. The sample processing time was also extended by the need to gauge sample device performance in terms of taxon abundance (as is standard in EA samples) rather than presence absence (as required by the indices NTAXA and ASPT).

In general, the sort times for sampling devices and sites reflected sample volume. The mean sort time required for airlift samples was the most consistent between sites and reflected the consistency in the volume and type of debris obtained (Figure 2). The medium naturalists' dredge produced small samples at one site, whilst the long-handled pond net provided samples of relatively small mean volume at 4 of the 6 sites. Compared to sample processing for standard assessments, the quantities of material collected with the medium naturalists' dredge and the airlift were not exceptionally large, but the high proportion of fine detritus found at some of the sites studied extended the sort times.

The rate at which new BMWP taxa were recovered during sample sorting and identification was compared between airlift, dredge and long-handled pond net. Mean recovery rates of BMWP taxa (NTAXA) per hour were: airlift - 2.06; medium naturalists' dredge - 2.14; long-handled pond net - 2.98. The airlift samples, though slower to sort, provided the most consistent return per hour.



	York Ouse	York Derwent	South Drove Drain	New Bedford River	Huntspill	Severn
airlift	8.2	7.5	8.3	8.6	9.1	8.1
dredge	10.2	2.6	7.2	10.7	10.2	9.5
long-handled pondnet	1	2.3	2.9	10	4.4	6.5

Figure 2. Comparison of mean sample sort times between sampler types and sites.

Table 4. Mean and standard deviation (SD) of NTAXA, BMWP Total Score and ASPT, by site for the four techniques tested. Additional replicates for different operators shown for the Airlift and Medium Naturalists' Dredge.

BMWP NTAXA

	Airlift		Dredge		LHPN		Margin	
	mean	**SD**	**mean**	**SD**	**mean**	**SD**	**mean**	**SD**
Y. Ouse	16.8	(1.2)	8.3	(2.1)	5.8	(1.7)	12.7	(2.2)
Y. Derwent	21.2	(0.7)	16.5	(2.9)	14.5	(2.2)	21.5	(2.9)
Y. Derwent 2	25.0	(2.6)						
Y. Derwent 3	22.3	(2.0)						
South Dr.	20.0	(1.2)	18.0	(2.7)	18.2	(0.9)	24.2	(3.0)
South Dr 2			19.8	(2.4)				
South Dr 3			18.8	(1.9)				
New Bedford	19.5	(1.9)	20.8	(3.4)	20.2	(1.7)	25.3	(2.7)
Huntspill	9.2	(1.9)	9.2	(1.1)	6.3	(0.6)	13.3	(3.0)
Severn	18.8	(2.0)	13.7	(6.0)	15.3	(5.0)	10.7	(3.9)

BMWP Total Score

	Airlift		Dredge		LHPN		Margin	
	mean	**SD**	**mean**	**SD**	**mean**	**SD**	**mean**	**SD**
Y. Ouse	75.6	(6.3)	33.3	(12.8)	16.3	(5.6)	53.7	(13.8)
Y. Derwent	128.0	(6.9)	90.8	(18.4)	84.7	(12.3)	115.5	(19.4)
Y. Derwent 2	149.7	(20.8)						
Y. Derwent 3	133.2	(8.6)						
South Dr.	87.5	(7.7)	78.3	(14.0)	81.8	(6.9)	110.3	(13.8)
South Dr 2			88.5	(10.8)				
South Dr 3			83.7	(11.5)				
New Bedford	98.5	(10.7)	100.7	(19.7)	99.8	(10.5)	126.8	(16.2)
Huntspill	34.2	(11.3)	31.8	(5.1)	21.5	(2.7)	51.7	(14.0)
Severn	97.8	(12.8)	65.5	(34.8)	77.7	(33.2)	45.8	(20.5)

ASPT

	Airlift		Dredge		LHPN		Margin	
	mean	**SD**	**mean**	**SD**	**mean**	**SD**	**mean**	**SD**
Y. Ouse	4.50	(0.09)	3.84	(0.78)	2.72	(0.37)	4.19	(0.41)
Y. Derwent	6.06	(0.21)	5.49	(0.36)	5.86	(0.30)	5.36	(0.31)
Y. Derwent 2	5.97	(0.30)						
Y. Derwent 3	5.98	(0.20)						
South Dr.	4.37	(0.18)	4.33	(0.18)	4.50	(0.22)	4.55	(0.09)
South Dr 2			4.46	(0.09)				
South Dr 3			4.39	(0.21)				
New Bedford	5.02	(0.14)	4.81	(0.24)	4.94	(0.14)	5.00	(0.17)
Huntspill	3.65	(0.47)	3.44	(0.18)	3.21	(0.09)	3.83	(0.23)
Severn	5.17	(0.21)	4.38	(1.19)	4.95	(0.55)	4.19	(0.48)

2.4.3 Biotic Indices

NTAXA

No one method yielded consistently higher NTAXA across all sites. Compared to the other techniques, the airlift produced samples with high NTAXA from the Yorkshire Ouse and Derwent, and the margin produced samples with high NTAXA from the South Drove Drain and New Bedford River (Table 4). When comparisons are restricted to the deep-water channel sampling methods the airlift produced higher NTAXA at all sites except the New Bedford River and River Huntspill.

BMWP Scores

No one method yielded consistently higher BMWP Scores across all sites (Table 4). The airlift samples from the Yorkshire Ouse and Derwent generated the highest BMWP. The mean BMWP Scores derived for each site confirm that the airlift sampler produced the highest BMWP Scores at 5 of the 6 sites, when comparisons are restricted to the deep-water sampling methods (Figure 3).

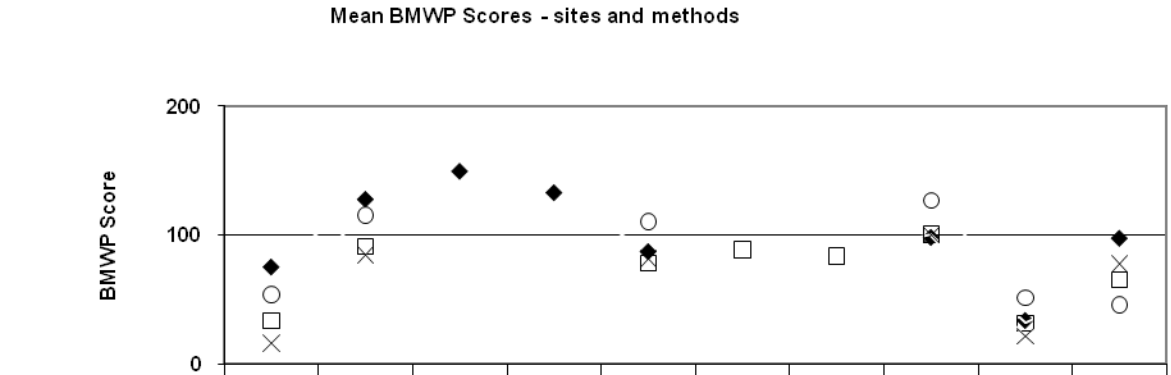
ASPT

The ASPT derived for each replicate sample generated similar trends to the BMWP Scores (Table 4). The airlift produced the most consistent ASPT within sites. The mean ASPTs derived for each site confirm that the airlift sampler also produced the highest ASPTs at 5 of the 6 sites, when comparisons are restricted to the deep-water sampling methods (Figure 4).

Taxon accretion rates

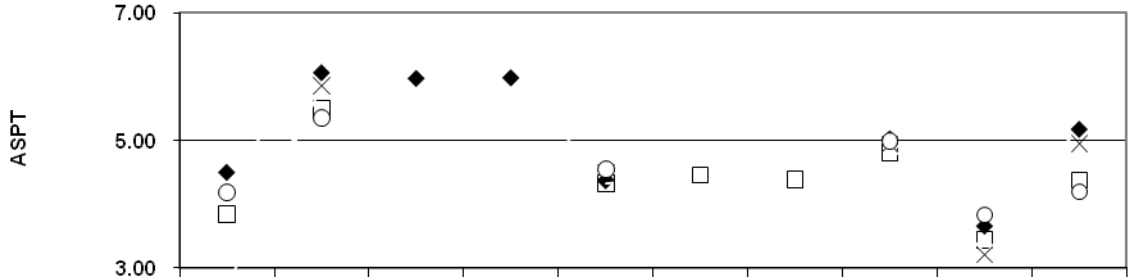
Rather than comparing the number of scoring taxa (NTAXA) as a single value, smoothed 'species' accretion curves were created for BMWP scoring taxa using the software package 'Species Diversity and Richness - Version 2' (PISCES Conservation Ltd, 1998) to determine the number of samples required sufficient to capture all the taxa present at the site that could eventually be captured by that sampling method. This approach highlighted the differing results generated by choice of sampling method between sites (Figure 5). For the Severn and New Bedford sites, sampling method had least influence on the total taxa recorded, or on accretion rates. Two sites (Huntspill and South Drove Drain) showed similar taxon recovery by airlift and medium naturalists' dredge, with relatively poor recovery rates by the long-handled pond net. The Yorkshire Ouse and Derwent displayed strongly contrasting taxon recovery and accretion rates between all methods. *The long-handled pond net produced the poorest total taxa count at four of the six sites:*

Sampling effort and yield were compared, in terms of the relationship between the calculated taxon accretion rate and numbers of animals recovered and identified. The standard RIVPACS sampling approach is designed to recover a minimum of 70% of the NTAXA present at a site without compromising site quality assessment. Bass *et al*. (2000) selected an 80% recovery rate of the maximum NTAXA recorded at each site for comparisons. The time required to achieve 80% recovery at each site was calculated by combining the known sort time for each sampling method, the number of samples and equivalent number of specimens requiring identification and counting (Table 5). It should be noted that the sample processing included estimations of taxon abundance as is now standard for RIVPACS samples.
The long-handled pond net could not achieve 80% of the available taxa at four of the six sites.



	Y.Ouse	Derwent 1	Derwent 2	Derwent 3	South Drove 1	South Drove 2	South Drove 3	New Bedford	Huntspill	Severn
◆ airlift	76	128	150	133	88			99	34	98
□ dredge	33	91			78	89	84	101	32	66
× long-h p-net	16	85			82			100	22	78
○ margin pondnet	54	116			110			127	52	46

Figure 3. Mean BMWP Score for each sampling method and site.



	Y.Ouse	Derwent 1	Derwent 2	Derwent 3	South Drove 1	South Drove 2	South Drove 3	New Bedford	Huntspill	Severn
◆ airlift	4.50	6.06	5.97	5.98	4.37			5.02	3.65	5.17
□ dredge	3.84	5.49			4.33	4.46	4.39	4.81	3.44	4.38
× long-h p-net	2.72	5.86			4.50			4.94	3.21	4.95
○ margin pondnet	4.19	5.36			4.55			5.00	3.83	4.19

Figure 4. Mean ASPT for each sampling method and site.

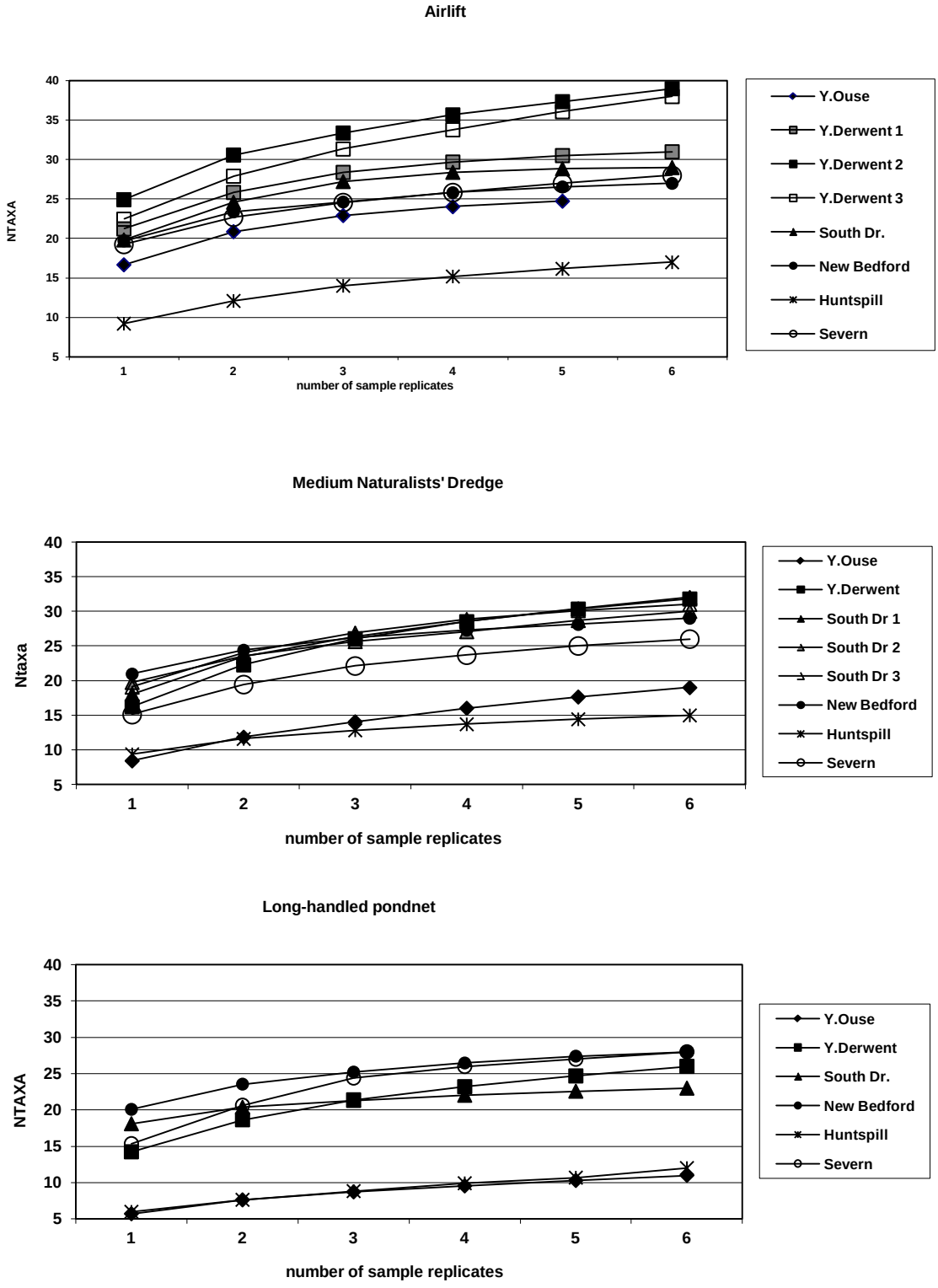


Figure 5. Smoothed taxon accretion curves indicating the predicted NTAXA found in any single, pair, 3, 4, 5, or 6 random samples (out of the total of six replicate samples taken by that method) at each site surveyed using the airlift, medium naturalists' dredge and long-handled pond net. Flattening of curves to a plateau indicates the maximum NTAXA retrievable with that method and the number of replicates required to achieve this.

Table 5. Comparison of time (hours) and the equivalent number of sample replicates required to recover 80% of the BMWP Scoring Taxa recorded at each site by the deep-water sampling methods tested. (Fastest options highlighted). Note variable results between BAMS series. N/A denotes the yield cannot reach 80% of the recorded taxa.

BMWP NTAXA

	Airlift			**Dredge**			**LHPN**		
	Samples to yield 80% taxa	Sort time per sample	sort time + Identification	Samples to yield 80% taxa	Sort time per sample	sort time + Identification	Samples to yield 80% taxa	Sort time per sample	sort time + Identification
Y. Ouse	2	8.2	**20.4**	6	10.2	73.2	N/A	N/A	N/A
Y. Derwent	4	6.2	32.8	5	2.6	21.2	N/A	N/A	N/A
Y. Derwent 2	2	8	**20**						
Y. Derwent 3	3	8.3	31						
South Dr.	3	8.3	30.9	3	5.2	**21.6**	N/A	N/A	N/A
South Dr 2				3	8.9	32.7			
South Dr 3				3	7.3	27.9			
New Bedford	2	8.6	**21.2**	2	10.7	25.4	2	10	24
Huntspill	3	9.1	**33.3**	4	10.2	48.8	N/A	N/A	N/A
Severn	2	8.1	**20.2**	3	9.5	34.5	3	6.5	25.5

2.4.4 Inter-operator differences

If the biological information obtained for a site is highly dependent on who took the sample, then it is more difficult to assess spatial and temporal changes when different personnel have been used. Therefore, it is important to assess the sampling variability between operators.

In their study, Bass *et al.* (2000) assessed differences in NTAXA, ASPT, BMWP Score and total number of individuals per sample attributable to operator. This was possible at two sites: at the Yorkshire Derwent site, three operators each took six replicate airlift samples, and at the South Drove Drain site, three operators each took six replicate medium naturalists' dredge samples. Both parametric (ANOVA) and non-parametric (Kruskal-Wallis ANOVA by ranks) tests were used. Inter-operator differences were not statistically significant for any index, for either the airlift or medium naturalists' dredge sampling method, but this may have been due to the small number of replicates and hence the low power of the test to identify differences. However, the estimates of the practical importance of inter-operator effects on total variance in index values, which was not biased by replicate or operator number, suggest that there is little or no inter-operator effect on ASPT values. For the airlift sampling method, the difference between operators may account for 20-30% of total replicate variation in both NTAXA and BMWP score (which are highly correlated). For the medium naturalists' dredge sampling method, difference between operators may account for 20-30% of total replicate variation in total number of individuals recovered. A more intensive replicated sampling study and analyses of uncertainty and inter-operator differences were undertaken in the NS share project (Neale *et al*. 2006; see Section 2.7).

2.4.5 Margin Pond Net Samples

Bass *et al.* (2000) also investigated the contribution of habitats at the watercourse margin to water quality status, and the distribution of BMWP taxa between the margins (both banks) and deep water habitats. The margin samples targeted the habitats accessible when using a standard FBA pond net (1.5 m handle). Each sample comprised three 1 minute sweeps (analysed separately and combined) from each bank. The samples did not incorporate any manual search.

There were differences in BMWP Score, NTAXA and ASPT between samples collected at the margins and the deep water, but the differences were neither consistently higher nor lower, nor of consistent size. There were consistent differences in BMWP Scores between the two banks at two sites (Huntspill and Severn) which were not evident for ASPT. The authors suggested that differences in index values may be caused by the abundance of macrophytes, although analysis of their data reveals this was not evident at all sites (Figure 6).

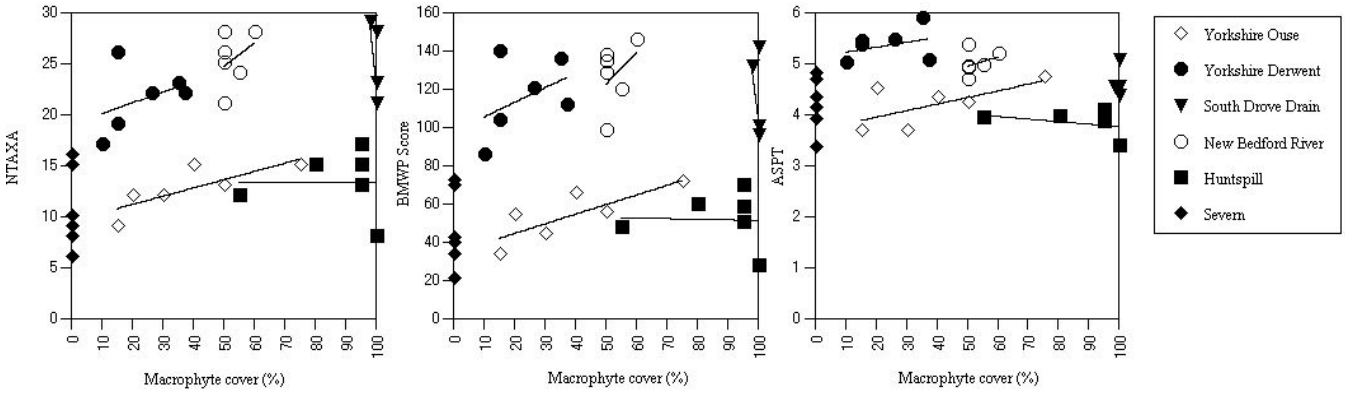


Figure 6. Relationship between macrophyte cover and index scores of 1 minute marginal sweep samples at six deep river sites.

2.4.6 Taxonomic composition

Deep-water sampling methods generally excluded taxa strongly associated with emergent vegetation and similar habitats confined to the watercourse margin. The contrasts in faunal composition were strongest between samples collected from the margin and the deep-water, rather than between deep-water methods. At sites where the medium naturalists' dredge passed through marginal vegetation at the end of its retrieval, some additional margin fauna were incorporated in the sample. Certain taxa were recovered exclusively from deep-water benthic samples (Table 6).

For all sites a combination of the results from deep-water and margin samples yielded higher NTAXA than samples from just one zone, although the results were not consistent. The combined airlift and margin samples yielded the highest NTAXA at three of the six sites and at the remaining three sites their totals were within one or two taxa of the site maximum obtained from combining medium naturalists' dredge plus margin, or long-handled pond net plus margin. The relative contribution from margin samples did not consistently mirror the level of habitat complexity at sites.

Bass *et al*. (2000) did not undertake a detailed investigation into the relationship between index scores from deep water and corresponding margin samples.

Table 6. Occurrence of taxa confined to deep-water samples (n - number of sample replicates, out of 18, in which the taxon was present).

Site	Deep-water	n
Yorkshire Ouse	Dendrocoelidae	5
	Planariidae	9
	Leptoceridae	3
	Simuliidae	1
Yorkshire Derwent	Hydropsychidae	7
	Unionidae	3
South Drove Drain	Unionidae	2
New Bedford River	No additions	
Huntspill	Unionidae	17
	Leptoceridae	3
Severn	Corophiidae	13
	Heptageniidae	5
	Ephemeridae	3
	Aphelocheiridae	2
	Elmidae	10
	Hydropsychidae	10
	Brachycentridae	6

Table 7. Comparison of the NTAXA recorded from deep-water samples, margin pond net samples and combined methods at each site. The combined methods yielding the highest NTAXA are highlighted.

Sampling method	NTAXA					
	Ouse	**Derwent**	**South Drove**	**New Bedford**	**Huntspill**	**Severn**
Margin pond net	24	34	37	36	23	25
Airlift 1	25	31	29	27	17	28
Dredge 1	19	33	30	29	15	25
Long-handled pond net	11	26	23	28	12	28
Combined airlift and margin pond net	31	38	40	36	28	35
Combined dredge and margin pond net	27	39	39	37	24	29
Combined long-handled pond net & margin pond net	27	37	39	38	26	36

2.4.7 Conclusions

The airlift yielded the highest mean number of taxa at four of the six sites, and the same number as the dredge at one site. The *long-handled pond net performed poorly*.

Accretion curves for the airlift flattened out after fewer replicates and at higher NTAXA and than for the dredge samples.

Some series of *long-handled pond net* samples also reached a taxon accretion plateau, but in these cases the *maximum achievable NTAXA was considerably lower than recovered by other sampling devices* at the same sites.

All accretion curves indicate that a single deep-water benthic sample taken from an area of 1.5 m^2 is not sufficient to recover 80% of the NTAXA recorded from each site. For routine monitoring the area sampled should be greater than 1.5 m^2.

In terms of BMWP taxon representation, the airlift sampler performed more effectively than the dredge at most sites and required fewer sample replicates to yield 80% of the NTAXA detected at each site.

The dredge yielded very similar results to the airlift at three sites, but only when all six sample replicates were taken into account.

The long-handled pond net under-performed in terms of recovering available BMWP Scoring Taxa and should be discounted as a reliable sampling method for deep-water benthos in wide rivers such as those studied here (i.e. wider than 15 m).

The time required to process samples was strongly influenced by sample volume and this reflected site conditions, the sampled area and method. The sample processing time was comparable among the different methods tested, and least variable for the airlift.

An estimate of the average time taken to process sufficient samples to recover 80% of taxa indicated that the air lift was most efficient, requiring only 2 or 3 samples and less time at nearly all sites. However, it was noted that there are differing costs of manpower, equipment and safety aspects of the particular sampling devices that were tested.

Subsequently, the Naturalists Dredge was recommended for routine monitoring work in exceptional circumstances only, on health and safety grounds (Rayson 2000).

2.5 Blocksom and Flotemersch (2005)

As part of the extension of the USA national bioassessment programme into non-wadeable rivers a series of tests were conducted by the US EPA to compare existing and refine deep water sampling methods. In a field based comparison, Blocksom and Flotemersch (2005) compared six techniques recommended by three organisations. The US Environmental Protection Agency's Environmental Monitoring and Assessment Program, Surface Waters techniques comprised either drift net or kick net methods (Lazorchak *et al.*, 2000). If current velocity was at least 0.05 m/s at the sampling point, two drift nets (30.5 cm × 45.7 cm opening) were deployed at the lower end of a site for 3–4 h, with one net in shallower and one in deeper water. The kick net method consisted of two 20 s kicks using a rectangular frame net (50 cm ×30 cm) at each of the 11 transects. All kicks were conducted on one bank and compiled into a single sample for the entire site. Two methods from the U.S. Geological Survey's National Water Quality Assessment Program were tested. The richest targeted habitat method consisted of sampling five to six 50 cm×50 cm areas of the richest habitats (e.g., rocks, large woody debris, macrophytes) along either bank of a 1000-m stretch of the reach. All samples were then compiled into a single sample for each site. Two methods recommended by the Ohio Environmental Protection Agency, Division of Surface Water Biocriteria were tested (Ohio EPA, 1987). The 500 m qualitative multihabitat method consisted of sampling all habitat types along both banks of a 1000 m section of the reach using a D-frame dip net for a minimum of 30 min or until no new taxa were observed by gross examination. In the artificial substrate method Hester-Dendy multi-plate substrate samplers were deployed at each site for approximately 6 weeks. Total surface area of each sampler was approximately 0.092 m^2. Upon retrieval, samplers were disassembled, and organisms and debris processed. All kick/sweep net samples were collected with nets of standard handle length restricting the depth of water from which a sample can be collected to *circa* 1m.

Forty-two metrics were used to assess the performance of the different approaches, and these correlated with measures of physical and chemical characteristics of the river reach and the land use in the riparian corridor.

The *drift nets were not effective.*

The *results from the Hester-Dendy artificial substrate samplers differed greatly from other sampling methods.*

Although metric values were similar across certain sampling methods, the metrics that significantly correlated with the abiotic variables differed among methods. At sites of deeper average depth there was more consistency among methods, probably due to limited access by wading at these sites, which limits the overall accessible sampling area in a reach: most of the methods tested relied upon access to the river margins by wading.

Despite the objective of this work being to test deep water sampling methods, the *methods tested by Blocksom and Flotemersch (2005) sampled shallow areas of deep rivers and largely left the deeper mid-channel habitats un-sampled*. This approach does not comply with that used in RIVPACS models, where all habitats are sampled in proportion to their occurrence. Furthermore, the results from such an approach may be influenced by the cross-sectional profile of the river; less of the river width will be sampled where rivers are deep to the margin.

2.6 Blocksom and Flotemersch (2008)

Following the publication of the USA EPA recommended large river macro-invertebrate bioassessment protocol (Flotemersch *et al.* 2006) an assessment of the precision and sensitivity of the method was published. Replicate samples, each comprising the identification of 300 random individuals collected by a kick sample covering 0.25 m^2, were collected at 19 sites on four rivers, spanning two depth classes (<4 m and >4 m thalweg depth). The replicate samples provided data for estimates of precision in the laboratory and field, and abiotic variables allowed for measurements of overall sensitivity. The metrics used were: total taxa richness, EPOT (Ephemeroptera, Plecoptera, Odonata and Trichoptera) taxa richness, percent tolerant individuals (taxa with generalized tolerance values >6; Klemm *et al.* 2003), percent Chironomidae and percent dominant taxon.

Precision and performance differed between the two depth classes of rivers, particularly the percentage of variance attributable to replicate samples (NB each sample comprised multiple sub-samples that were amalgamated). When compared to the precision of RIVPACS on similar metrics measured as the percentage of variance attributable to individual replicate samples (Clarke *et al.* 2006) the USA EPA methodology appears to perform similarly. It should be noted that as these methods are based on samples from the shallow (< 1 m) parts of deep rivers, a smaller proportion of the river bed is available for sampling at the deeper sites which will influence the variance between replicate samples (Table 8). When the percentage total within site variance of the USA EPA large river methodology is compared to the most precise technique tested by Neale *et al.* (2006), the US EPA method performs similarly (Table 8). It should be noted that the EPA methodology tested samples each collected from approximately 1/6 of the benthic area sampled by Neale *et al.* (2006), and restricted to the first 300 individuals identified: the small sample area and restricted count tends to reduce variance among samples.

Table 8. The percentage of total variance attributable to replicate samples and total within site variance of the US EPA large river methodology for shallow (thalweg depth < 4 m) and deep (thalweg depth > 4 m) river sites compared to measures for standard RIVPACS kick samples as tested by Clarke et al. (2006) and the most precise deep river method tested by Neale et al. (2006). Note different metrics are used in the US and UK studies.

Metric	% Replicate variance			% Total within site variance		
	Shallow sites	Deep sites	RIVPACS	Shallow sites	Deep sites	Airlift
Total Taxa	6.4	13.2	8	39.1	30.6	24
Number of families			6			
BMWP Score						19
ASPT			9			23
EPOT Taxa	0.5	5.5		44.2	24.7	
EPT Taxa			5			
% Tolerant individuals	14.2	1.9		19.7	7	
% Chironomidae	22.3	16.2		56.1	19.6	
% Dominant taxa	21.6	13.5		45.7	17	
% Gatherers/collectors			4			
% Oligochaeta			7			

2.7 Neale, Kneebone, Bass, Blackburn, Clarke, Corbin, Davy-Bowker, Gunn, Furse and Jones (2006)

This report, undertaken as part of the INTERREG North South Shared Aquatic Resource (NS Share) project, provided a comprehensive field based assessment of four deep water techniques, samples collected from the margin with a standard pond net, or from the channel with a long-handled pond net, the light naturalists' dredge or an airlift (see Figure 1). Replicate samples were collected from thirteen sites covering a wide range of conditions, covering coarse to fine substrates in both natural and artificially deepened channels.

The techniques were compared in terms of:

- ease of use,
- time taken to process the sample,
- biotic indices,
- uncertainty,
- cost-effective precision,
- influence of the environment on performance,
- community composition, and
- comparability to a standard kick sample.

A survey of the opinion of the workers involved in collecting the samples was conducted indicating that all the techniques are equally difficult to use in the field, and the only factor that discriminated between the techniques was the requirement of a boat to collect an effective sample. The airlift always requires a boat to collect a sample: however, a boat was also required to collect a sample from the margin of narrow, deep rivers as they tend to be engineered or deeply incised and have steep banks. The use of a boat has clear implications for the time and manpower required to collect a sample, together with health and safety implications.

There were differences among the techniques in the amount of time it took to sort and process the samples, with the airlift, which often produced large samples, taking longer on average to sort and hence process than the other three techniques (Figure 7).

On face value it appeared that the light naturalists' dredge and long-handled pond net are the most efficient techniques in terms of the time and effort required to collect and process each sample, and the airlift the costly in terms of time and effort per sample. However, it should be noted that Neale *et al.* (2006) undertook further analysis incorporating uncertainty to determine cost effective precision which gave a very different perception of efficiency (see below).

Table 9. Environmental Conditions at the Sites used in the NS Share Project.

Velocity category (1 = <10 cm s^{-1}, 2 = 10-25 cm s^{-1}, 3 = 25-50 cm s^{-1}, 50-100 cm s^{-1}, >100 cm s^{-1}), B/C = % boulders/cobbles, P/G = % pebbles/gravel, S = % sand, S/C = % silt/clay, M = % macrophyte cover

River	Site	Width (m)	Depth (m)	Velocity	B/C	P/G	S	S/C	M
Blackwater	Blackwatertown	25	2	1		20	60	20	2.5
Blackwater	Moy	38	3	1		50		50	5
Clogh	Glarryford	10	1.5	1		10	70	20	
Erne	Rosscor Bridge	82	4.5	1	2	80			0.05
Finn	Wattle Bridge	25	2.5	1	5	5		90	10
Garavogue	Lough Gill	100	4	1			10	90	10
Leannan	Lough Fern	21	2	3	1	80	10		80
Main	Dundermot	14	1.5	1	5	50		45	
Moy	Arran Bridge	48	140	2	5	80		15	25
Owencarrow	New Bridge	12	2	1	10		70	20	30
Shannon	Hartley Bridge	42	3.5	1		22	33	45	10
Sillees	Carr Bridge	15	1.5	2					5
Strule	Abercorn Bridge	31	1.5	2	90	5	5		

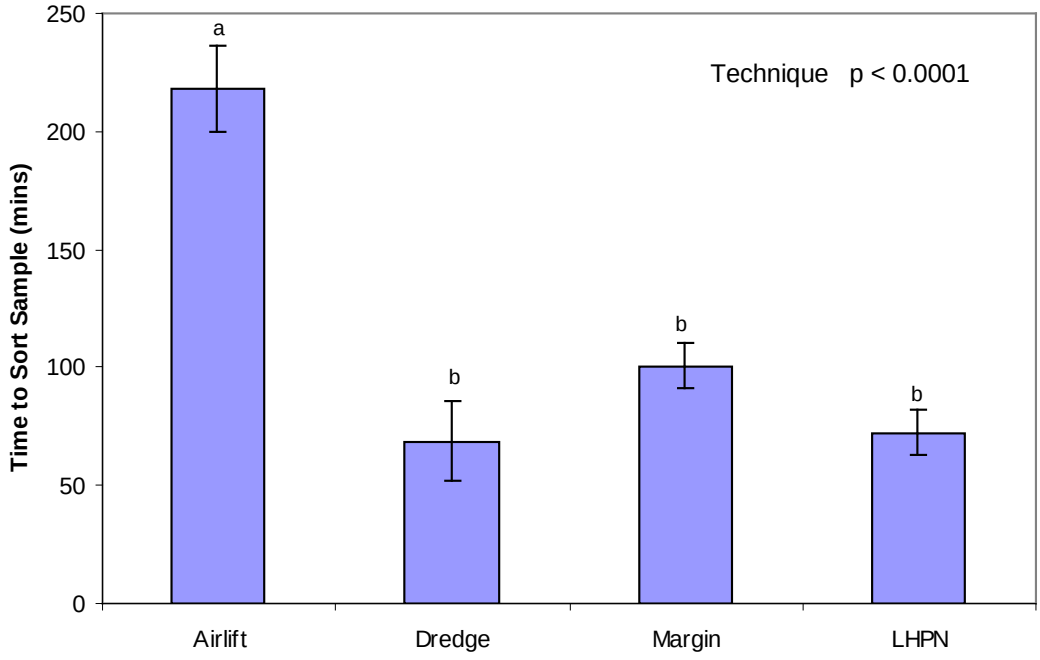


Figure 7. Influence of technique on the time taken to sort the samples collected with the four deep water techniques tested. Mean values shown ±SE. Different letters indicate significant differences among mean values as identified by Tukey's test, shared letters indicate no significant difference.

The light naturalists' dredge and long-handled pond net performed poorly relative to the airlift, in terms of all the metrics used to assess the macroinvertebrate community (NTAXA, BMWP Score, ASPT:

Figure 8), particularly when the substrate comprised a high proportion of boulders. The light naturalists' dredge and long-handled pond net also performed poorly compared to the margin samples in terms of BMWP score and NTAXA. The margin samples had the highest NTAXA, and equal highest total BMWP score with the airlift. Importantly, the airlift had a significantly higher ASPT than all the other techniques, indicating that this technique sampled the more sensitive, high scoring taxa more effectively. It is important that the sensitive taxa are sampled, as these taxa will show most rapid response to changes in water quality.

Samples collected at the margin capture different components of the fauna to those collected from the river channel. There was a marked difference in taxonomic composition between the samples collected at the margin and those collected from the river channel, found in both the field trial and in the comparison with historic EHS data. Furthermore, there was little correlation between the metric scores of the samples collected from the margin and those collected from the river channel (Figures 9-11). The fauna of the margins seems to be responding to different pressures to the fauna of the river channel. Perhaps this was because the fauna that characterised the samples collected from the margin tend to live at the air-water interface or be associated with vegetation. Such taxa may be less sensitive to in-stream influences and be determined more by habitat structure at the margin. *The use of samples collected from the margins alone is not sufficient to describe site condition.*

The difference between the river channel and the margin increased with increasing width of the river channel, in terms of ASPT of the airlift and the margin samples, indicating an increasing divergence in the community sampled (Figure 12). It is possible that in narrow rivers representatives of the mid-channel fauna are present in the margins and caught by both techniques, whereas in wider rivers there is more spatial segregation between the margins and channel, together with the occurrence of sensitive, deep river taxa in the channel. Separation of the two techniques appears to occur below 20 m width. Mid-channel depth does not appear to have a significant influence.

It is not necessary that the "highest score" should determine which technique is recommended provided that a model is developed using a standard technique that samples a subset of the fauna that is sensitive to the pressure of interest. However, it should be noted that the sample from the margin sampled different components of the community, which is less sensitive (in terms of ASPT) and responding to different drivers to the samples from the river channel.

The light naturalists' dredge and long-handled pond net appeared to sample a subset of the taxa collected by the airlift, being characterised by a lack of (obligate) deep water taxa. Neither of these techniques were effective at providing an adequate sample of the macroinvertebrate fauna present. *The airlift provided the most adequate sample of the river channel fauna.*

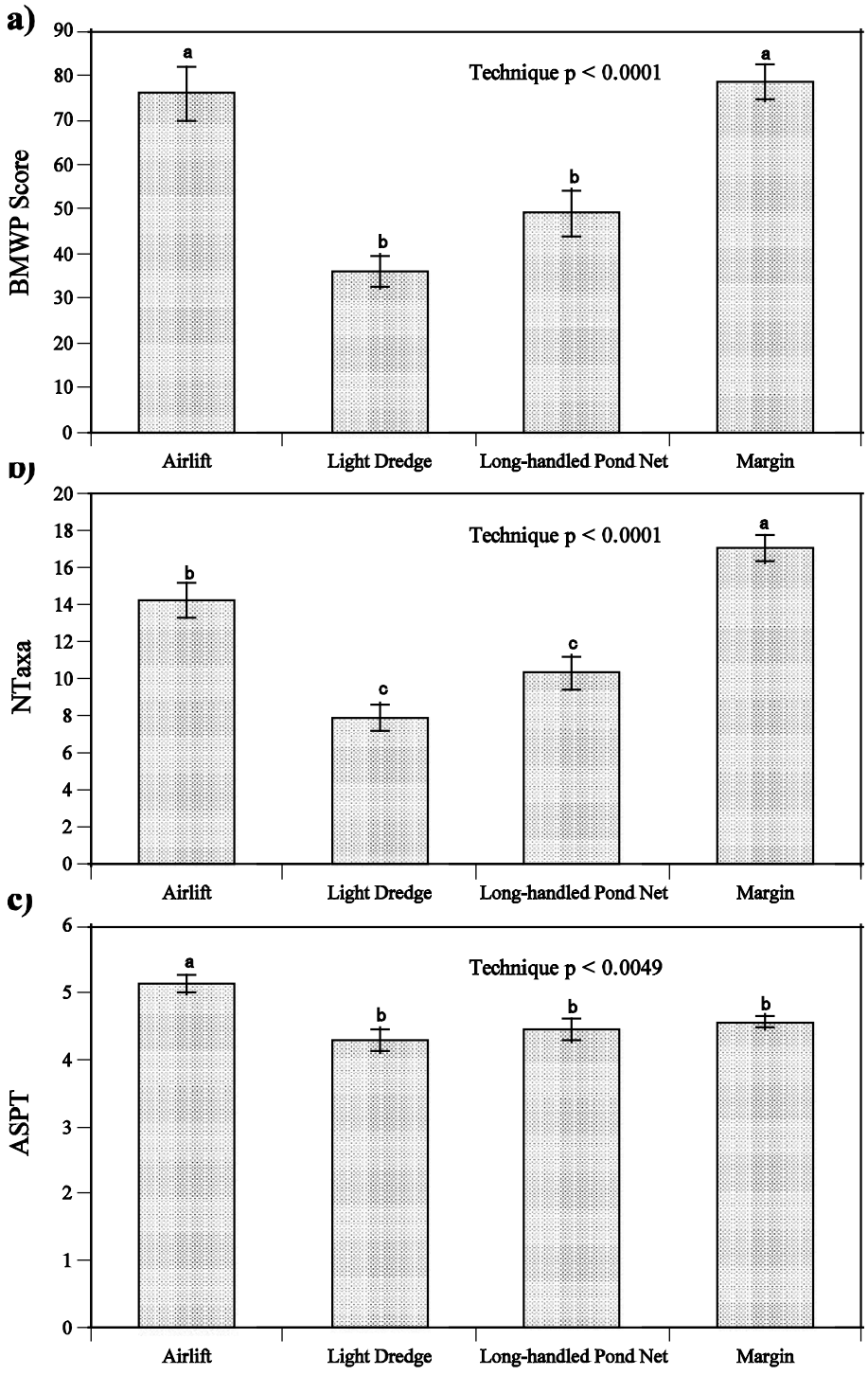


Figure 8. Influence of technique and site on a) total BMWP score b) NTAXA and c) ASPT of the samples collected with the four deep water techniques tested. Mean values shown ±SE. Different letters indicate significant differences among mean values as identified by Tukey's test, shared letters indicate no significant difference.

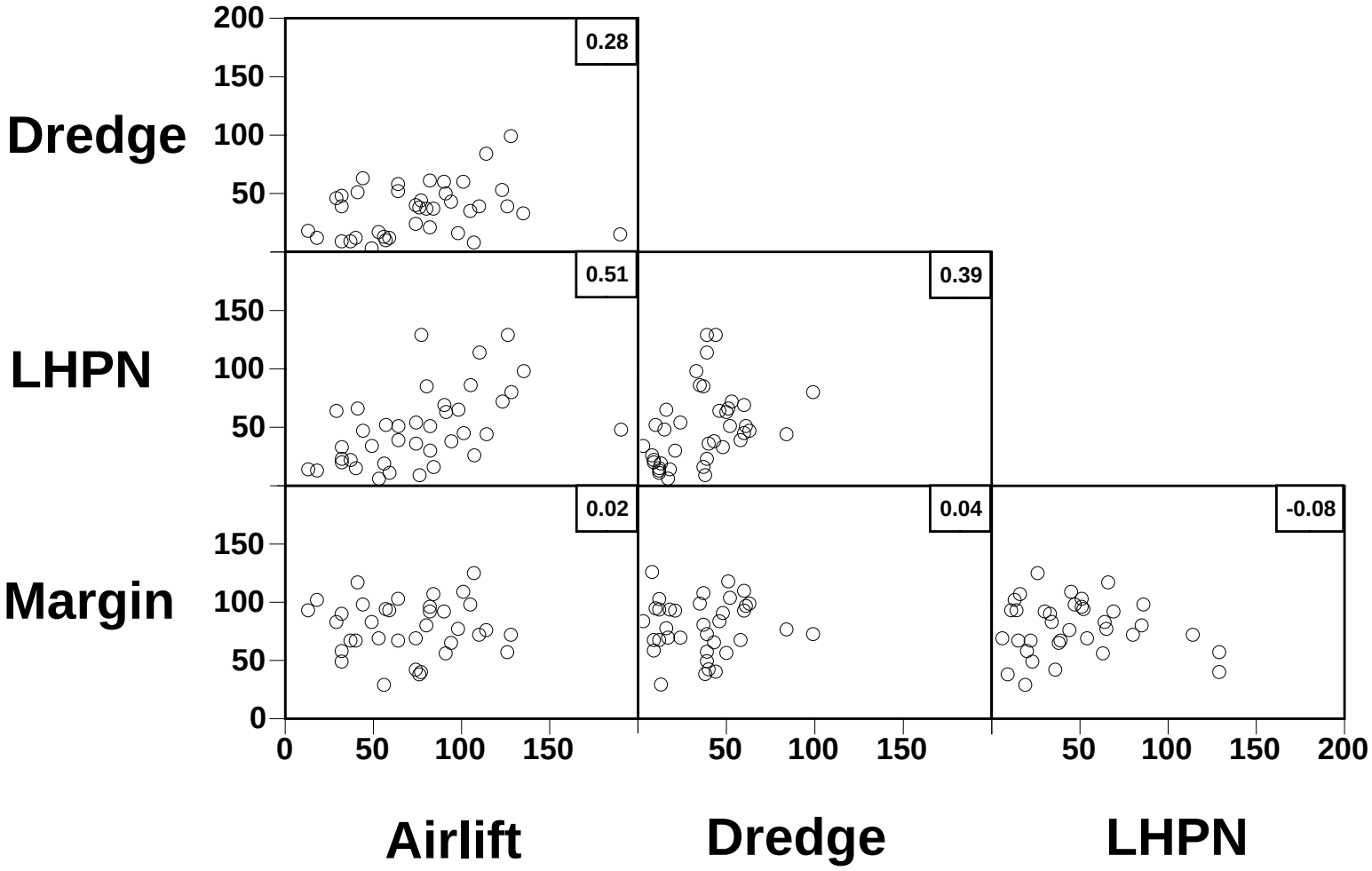


Figure 9. Matrix showing correlation between BMWP scores of the four deep water techniques, using pairs of matched replicates from the same site reach. R is shown in the top right hand corner for each combination.

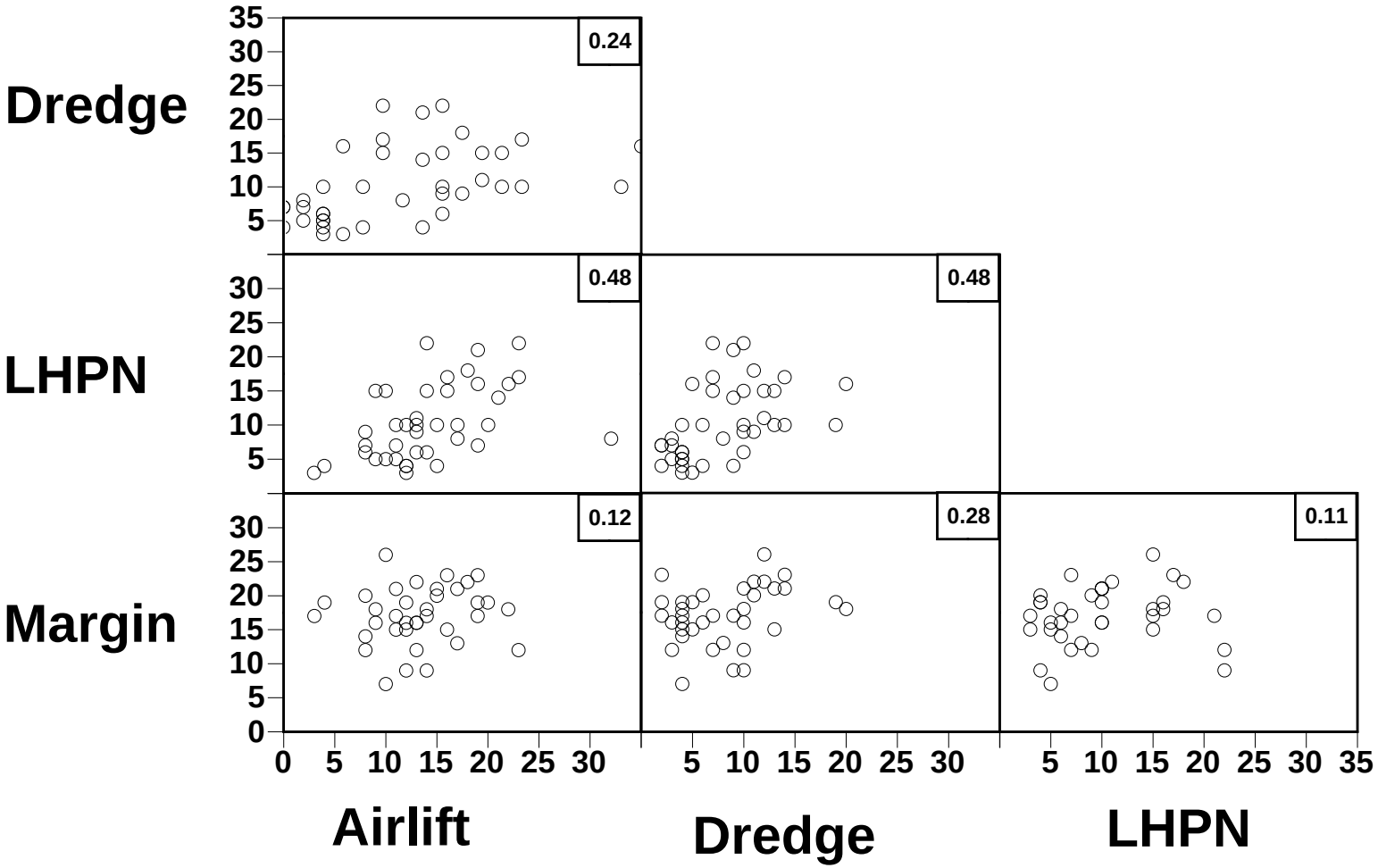


Figure 10. Matrix showing correlation between NTAXA of the four deep water techniques, using pairs of matched replicates from the same site reach. R is shown in the top right hand corner for each combination.

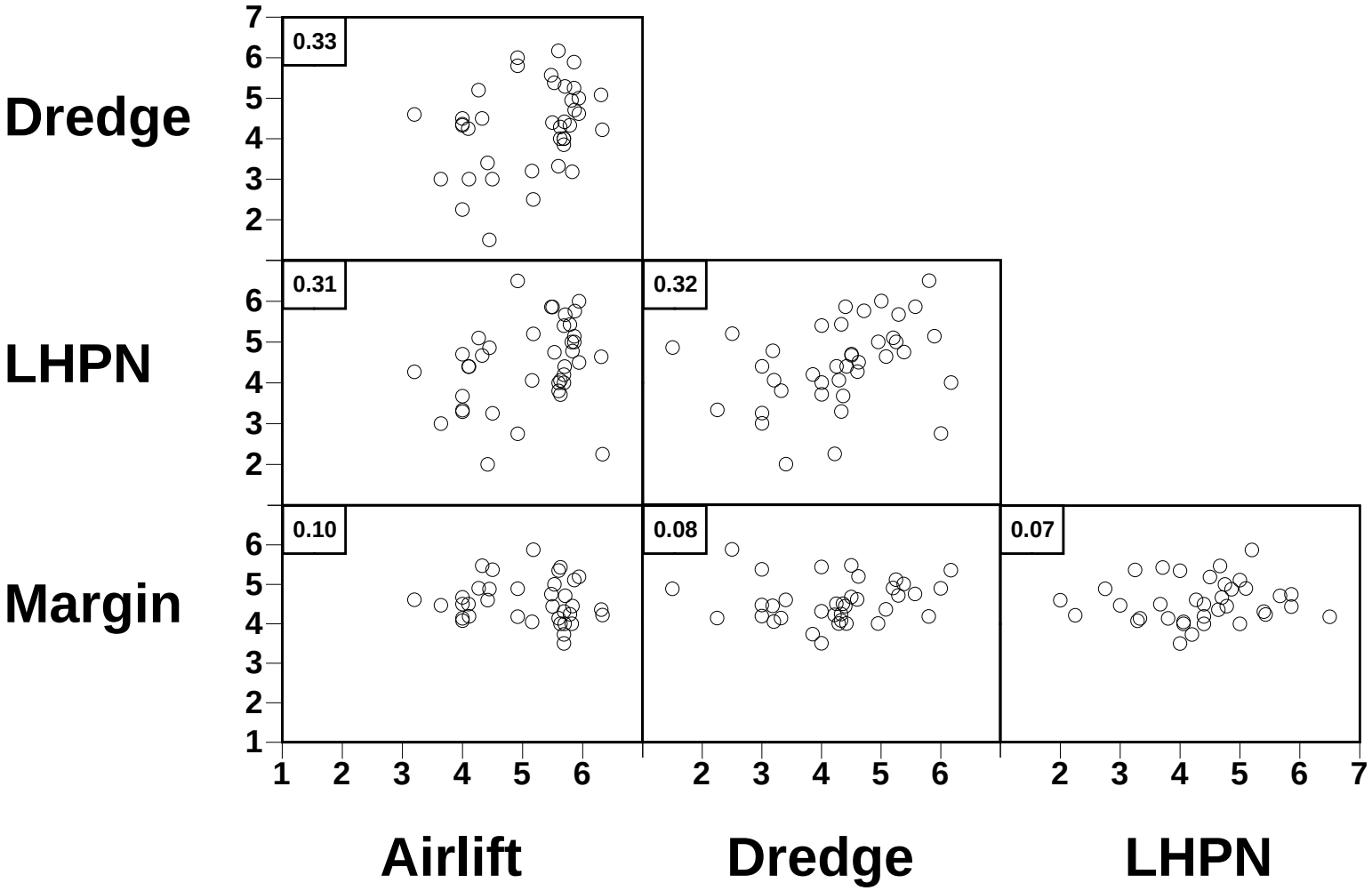
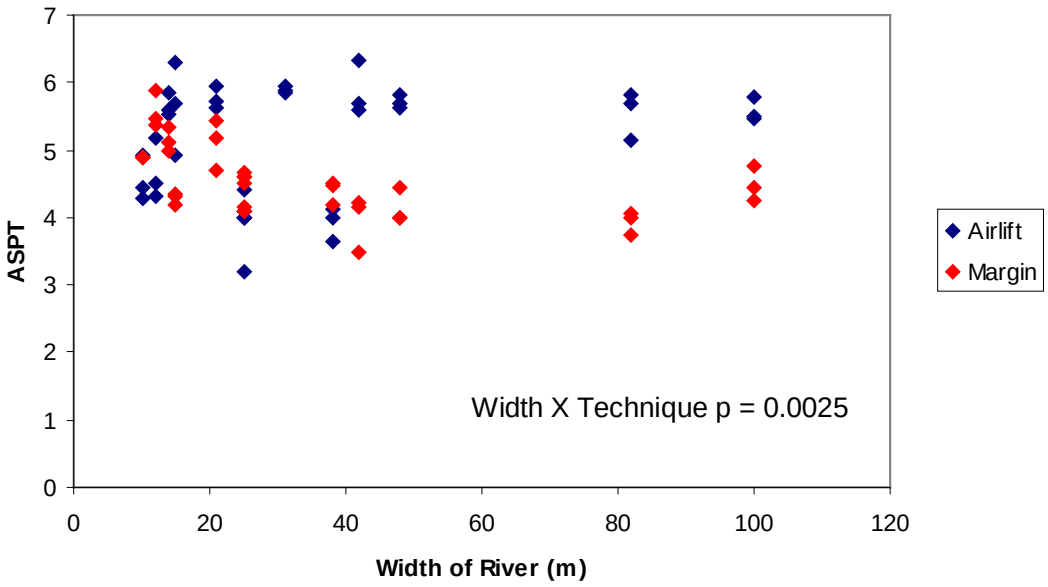


Figure 11. Matrix showing correlation between ASPT of the four deep water techniques, using pairs of matched replicates from the same site reach. R is shown in the top left hand corner for each combination.

a)



b)

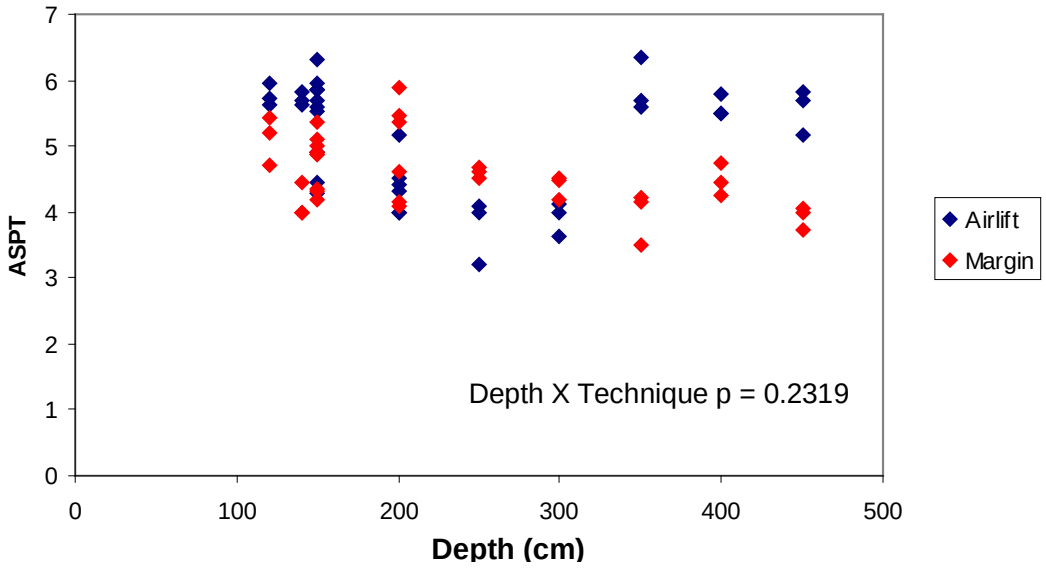


Figure 12. Influence of a) the width of the river channel and b) the depth of the centre of the river channel on relative performance of airlift and margin techniques in terms of ASPT. Results of interaction between technique and width and depth from Ancova shown; this interaction indicates differences in relative performance if significant.

For a sampling technique to be effective, sampling variability within a site and time period needs to be small relative to the real differences between sites in their biota and in their values for key biotic indices. In other words, samples should be repeatable within a site and discriminate between sites as far as possible. This has implications for detection of change and hence the number of samples that must be taken to achieve an adequate confidence of change, with consequences for the time and effort required to make a sufficiently precise assessment.

The airlift was the most precise of the techniques tested. More than 75% of the total variance was due to differences among sites for all three key metrics tested; implying that less than 25% was due to within site sampling variation (Table 10). Furthermore, there was little/no effect of operator (0-4% of total variance). Within site variance was so great among the samples collected with the light naturalists' dredge that little, if any, of the variance could be attributed to differences among sites (0% BMWP, 23% NTAXA, 5% ASPT). Operator effects accounted for a substantial and significant proportion of the variance for BMWP (64%) and NTAXA (42%). *The light naturalists' dredge has such high sampling variance and low repeatability that it is effectively useless for assessing and discriminating the ecological status of sites and need not be considered further.*

For the samples collected at the margin and those collected with the long-handled pond net, between 40% and 81% of the total variance in the key metrics tested across the study sites was attributable to differences between sites; implying that between 19% and 60% was due to within site sampling variation (Table 10).

In terms of precision, the airlift outperformed all the other techniques.

Sampling precision has implications for confidence of class. The technique with the lowest within site sampling variation will have the greatest confidence of class.

For a technique to be suitable for use in deep rivers it should be cost effective. The airlift was shown to have the lowest 'percentage within-site sampling variance' and thus the highest statistical precision and repeatability of results amongst the four techniques assessed. However, a single airlift sample has been shown to take a longer time to sort and process and, therefore, costs more per sample. Neale *et al*. (2006) combined these two aspects, namely sampling precision and sample processing time costs, to estimate the relative cost effectiveness of each technique.

Based on calculations of cost effectiveness for each technique and index, it was evident that *the increased costs in processing each airlift sample are outweighed by increased precision.* To achieve a sampling variance of less than 20% across all three metrics tested (BMWP, NTAXA, ASPT) will take an estimated average of 534 minutes for the airlift, compared to 735 minutes for the margin sample and 714 minutes for the long-handled pond net (Table 11). The airlift has the second most cost efficient precision for each of the individual metrics. It should be noted that the airlift could achieve a sampling variance of less than 20% in all three metrics (BMWP, NTAXA, ASPT) with just 2 samples indicating that this level of sampling variance could be achieved at a site in one year under the current programme of spring and autumn sampling: the other techniques would require samples collected over 3 (margin = 5 samples) and 4 years (long-handled pond net = 7 samples) to achieve this level of precision.

Obviously these comparisons still ignore any differences in costs associated with collecting the samples in the field. They also ignore any differences in the metric values achieved with the different techniques; it has been shown previously that for the airlift taxon accretion curves flattened out after fewer replicates than for the long-handled pond net and at higher number of taxa (Bass *et al.* 2000).

Given the higher precision of samples collected with the airlift (compared to other techniques) any increased costs of collecting samples using this technique could be counterbalanced by a reduced frequency of sample collection.

*Table 10. Estimates of sources of variance in BMWP Score, NTAXA and ASPT for each of the field sampling techniques (airlift, dredge, margin and LHPN). *, ** and *** denote site or operator variance component was statistically significant in ANOVA tests at the 0.05, 0.01 and 0.001 test probability level.*

		Technique			
	Variance	Airlift	Dredge	Margin	LHPN
BMWP	%Site	81 ***	0	60 **	68 **
	%Operator	4	64 *	10	13
	%Replicate	15	36	30	19
	% Within-site	19	100	40	32
NTAXA	%Site	76 ***	23	46 *	64 **
	%Operator	0	42 *	28	17
	%Replicate	24	35	26	19
	% Within-site	24	77	54	36
ASPT	%Site	77 ***	5	81 ***	40 **
	%Operator	0	0	0	0
	%Replicate	23	95	19	60
	% Within-site	23	95	19	60

Table 11. Comparison of the field sampling techniques (airlift, dredge, margin and LHPN) for sampling processing cost (time in minutes; number of samples shown in brackets) to achieve a sampling variance of less than Q% (20% or 10%) of the total variance amongst all sites in terms of BMWP Score, NTAXA, ASPT, and all 3 metrics. σ_I^2 and σ_W^2 denote between- and within- site variance estimates.

	Q%	Technique			
		Airlift	Dredge	Margin	LHPN
Per sample (mins)		267	94	147	102
(a) BMWP					
σ_I^2		1349	0	335	692
σ_W^2		308	551	226	329
	20%	267 (1)	(>100)	441 (3)	204 (2)
	10%	801 (3)	(>100)	1029 (7)	510 (5)
(b) NTAXA					
σ_I^2		29.05	5.18	8.5	19.2
σ_W^2		8.94	17.27	9.81	10.65
	20%	534 (2)	1316 (14)	735 (5)	306 (3)
	10%	801 (3)	2914 (31)	1617 (11)	612 (6)
(c)ASPT					
σ_I^2		0.616	0.062	0.264	0.495
σ_W^2		0.182	1.093	0.061	0.752
	20%	534 (2)	6674 (71)	147 (1)	714 (7)
	10%	801 (3)	15040 (160)	441 (3)	1428 (14)
Time to achieve precision in all 3 metrics	20%	534 (2)	(>100)	735 (5)	714 (7)
	10%	801 (3)	(>100)	1617 (11)	1428 (14)

For a technique to be suitable for use in deep rivers it should ideally be comparable to the methodology used for shallow rivers. There are several benefits in having this comparability.

1. It will not be necessary to develop independent models for deep rivers as they can be integrated into shallow water models.
2. A necessarily subjective, stepped boundary between deep and shallow rivers is avoided, such that the categorisation of a site as deep or shallow, in terms of the sampling technique to be used, will not influence the ecological status of the site.
3. Deep water reference sites can be classified along with shallow water reference sites, potentially reducing the number of deep water reference sites required for a RIVPACS-type model.

Two comparisons were made between the techniques tested here for sampling deep rivers and the techniques currently in place. The first was a direct comparison between a "standard kick sample" incorporated into the sampling strategy where conditions allowed, and the second was a comparison with historic data collected by EHS.

In terms of the time to process the samples, faunal composition and key metrics the samples collected with the airlift were the most similar to the standard RIVPACS kick sample. The light naturalists' dredge and long-handled pond net collected a similar fauna to the kick and airlift, but less effectively, missing parts of the fauna and producing smaller, and lower scoring samples. The samples from the margin sampled a different fauna to the kick and airlift. However it should be noted that this comparison was only undertaken at two sites. For full confidence of a lack of influence of technique on assessed biological quality a more thorough exercise would need to be undertaken.

The samples collected for this field study compared well with those collected by EHS as part of their routine monitoring. Most of the differences were the result of CEH only sampling in one season, and possibly by sampling in a smaller area than would be standard practice for EHS. EHS also combined marginal and kick/long-handled pond net samples. Nevertheless, there was segregation of samples collected from the margin (either on their own or in combination with kick/long-handled pond net samples) and samples collected within the river channel, irrespective of who collected them. There was no clear segregation of any technique tested from those used by EHS, so it was not possible to use this criterion to identify a technique that should be excluded from consideration.

There are obvious implications with respect to health and safety of the collection of any sample from deep water. Neale *et al.* (2006) were not qualified or tasked to assess these in detail, but were aware that they potentially include one or more of the following risks (applicable techniques given in brackets);

- Steep banks (margin, light naturalists' dredge, long-handled pond net).
- Over-stretching (margin, long-handled pond net).
- Carrying a boat to and from the site (airlift, margin from boat)
- Gaining access to the water with a boat (airlift, margin from boat).
- The use of a boat (airlift, margin from boat).
- The use of equipment from a boat (airlift, margin from boat).
- Carrying heavy equipment (airlift).
- Use of compressed air (airlift).
- Throwing heavy objects (dredge).

These risks and others will have to be assessed and taken into consideration when designing a monitoring strategy for deep rivers.

2.7.1 Conclusions

The airlift was recommended as the most suitable, precise, and cost-effective, technique for sampling deep rivers. The samples collected with the airlift were most comparable to a standard RIVPACS kick sample.

The airlift could achieve a sampling variance of less than 20% in all three metrics (BMWP, NTAXA, ASPT) with just 2 samples indicating that this level of precision could be achieved at a site in one year under the current programme of spring and autumn sampling: the other techniques would require samples collected over 3 or more years to achieve this level of precision. Hence, *increased costs of sample collection could be counterbalanced by reduced frequency of sampling*.

The light naturalists' dredge has no power to detect differences among sites, implying that it also cannot detect change within sites. Thus, it is effectively useless for the purpose of bioassessment. On this fact alone it was *recommended that the light naturalists' dredge is not considered for routine monitoring*. The poor performance of the light naturalists' dredge does not reflect on other dredges, but use of heavier dredges is restricted on health and safety grounds.

The long-handled pond net was considerably more variable, less representative and lower scoring than the airlift, often being close to the measures obtained with the light naturalists' dredge. The long-handled pond net did not sample obligate deepwater taxa effectively, tended to miss sensitive taxa, and did not perform well when the substrate was coarse. The long-handled pond net was not comparable to a standard RIVPACS kick sample.

Samples collected from the margins of deep rivers differ from those collected in the main channel. Wide rivers cannot be effectively sampled at the margin alone as the high scoring mid-channel fauna are overlooked.

3 Conclusions

For the sampling of deep rivers a method should be developed that is suitable for incorporation into routine monitoring, permits the sampling of at least one site per day, is scientifically defensible and statistically robust, and fits within the resource, and health and safety requirements of the agencies.

Aspects of the performance and suitability of the available techniques for sampling deep rivers have been rigorously tested under a wide range of environmental conditions encompassing many of the deep river types found in the UK and Republic of Ireland. Unless the design of equipment is modified from that used in these tests, it is recommended that there is no need for further comparative testing of deep river sampling methods. However, a more comprehensive study of comparability of the selected deep water sampling methods and a standard RIVPACS kick sample would need to be done, to ensure that the selection of sampling methodology does not influence measures of ecological quality.

Neale et al. (2006) suggested that the light naturalists' dredge has no power to detect differences in quality and thus is effectively useless. On this fact alone it is *recommended that the light naturalists' dredge is not considered for routine monitoring.* Heavier dredges (e.g. medium naturalist's dredge) perform far better, but use by throwing from the bank has been discounted on health and safety grounds (Rayson 2000): use of heavier dredges by towing from a boat may be possible but will require full field testing before any conclusions can be drawn.

Both Neale et al. (2006) and Bass et al. (2000) indicate that the long-handled pond net does not provide an adequate sample of the taxa present. This technique is also relatively imprecise. As a distinct mid-channel community appears to be present in wider deep rivers (> 20 m), following Benjamin (1999), it is *recommended that use of the long-handled pond net is restricted to narrow deep water courses* (< 15 m wide, ditches, etc). It has been suggested that a modified long-handled pond net, with the net angled to the shaft may perform better than the standard model (*pers. comm.* John Lucy of Republic of Ireland EPA) but this would require full field testing before any conclusions could be drawn, and does not in itself address the issue of sampling the mid-channel habitat.

The Environmental Protection Agency of the USA have adopted a strategy of sampling the shallow margins (<1 m depth) of large (deep) rivers. Whilst considerable testing of methods following this strategy has been undertaken (Blocksom & Flotemersch, 2005, 2008; Flotemersch *et al.* 2006) these methods have not been tested against methods that sample the deeper parts of the channel. Neale *et al.* (2006) indicated that *marginal samples do not correlate with mid-channel samples, suggesting that these two habitats respond to pressures differently*. Furthermore, the RIVPACS methodology developed for shallow river sites is based on a whole river channel assessment, sampling the available habitats in proportion to their occurrence: the US EPA methodology leaves large sections of the available habitat (everything > 1 m deep) un-sampled. Channel cross-sectional profile appears to influence both uncertainty and representativeness of samples collected following this strategy. It is *recommended that a strategy for routine monitoring is developed that samples both mid-channel and marginal habitats*.

The work of Neale *et al*. (2006) has implications for the development of sampling strategies for other water body types that only sample marginal vegetation: such tools will not represent the pressures on the open water adequately. It is also recommended that agencies replace monitoring activities based on sampling accessible areas of deep rivers with methods that provide a sample that represents all habitats, shallow and deep, as soon as is feasibly possible.

The airlift provides better representation, more sensitivity and precision, and is more cost-effective (in terms of processing samples) than other methods routinely used for monitoring. It is *recommended that a strategy for reference sample collection and routine monitoring of deep rivers is developed where mid-channel samples are collected with an airlift.* The design of such an airlift should follow that which has been tested. The design of such a strategy

will depend on being able to undertake required activities in a safe and cost effective manner given the available resource. In a following report (Davy-Bowker, Jones & Murphy 2012), this project will present the results of an ergonomic assessment undertaken on the use of the airlift for routine monitoring of deep water courses.

4 Recommendations

Based on the evidence compiled in this review it is recommended that:

1. A specific and standardised deep water technique is used in all sites where a large proportion (provisionally > 40%) of the width is too deep to safely wade. [In practice we suggest this depth is approximately 80 cm]
2. Samples should not be collected from shallow patches of deep rivers [as this violates the assumptions of RIVPACS and is prone to the impacts of natural changes in geomorphology over time.]
3. The Light Naturalists' Dredge, which is currently recommended, should not be used [as this technique is unable to confidently detect differences in quality (Neale *et al.* 2006).]
4. For narrow deep water courses (provisionally less than 15 m average width subject to verification later in this project) samples should be collected with a long-handled pond net. [Based on recommendations of Benjamin (1998). In wider water courses the channel and marginal communities segregate (Neale *et al.* 2006) and the long-handled pond net consistently fails to capture the full complement of available taxa (Bass *et al.* 2000, Neale *et al.* 2006), resulting in low index scores and hence lower assessments of quality.]
5. For wide deep water courses (provisionally greater than 15 m average width), samples from the channel should be collected with a Yorkshire pattern airlift. [Based on the recommendations of Bass *et al.* (2000) and Neale *et al.* (2006). Airlift samples are the most precise, most cost-effective, and consistently capture a wide range of available taxa, particularly sensitive taxa, resulting in high index scores and better assessments of quality (Neale *et al.* 2006). Airlift samples are the most similar to those collected using the standard kick net procedure of shallow waters (Neale *et al.* 2006), although this may need to be verified across different river types.]
6. A sample from the margin (equivalent to the 1 minute search of the shallow water technique) should be combined with a sample from the channel collected with either an airlift or a long-handled pond net. [Based on Bass *et al.* (2000), indicating that the widest range of taxa is achieved by combined samples, and Neale et al. (2006), indicating that margin and channel fauna respond to different stressors.]
7. The high precision of Yorkshire pattern airlift samples presents an opportunity to counterbalance the increased costs of sample collection with a reduced sampling frequency for deep rivers. [Based on Neale *et al.* (2006).]

References

ÅAREFJORD J. (1972) The use of an air-lift in freshwater bottom sampling. A comparison with the Ekman bottom sampler. Verhandlungen der Internationale Vereingung für theoretische und angewandte Limnologie, 18, 701-705.

ANDRE P., LEGENDRE P. & HARPER P.P. (1981) La sélectivité de trios engins d'échantillonnage du benthos lacustre. (The selectivity of three samplers for lake benthos – in French). Annales Limnologique, 17, 25-40.

BAKANOV A.I. (1979) New bottom grab designs and assessment of the aggregation state of the benthos. Hydrobiological Journal, 15, 77-82.

BAKER J.H., KIMBALL K.T. & BEDINGER C.A.Jr. (1977) Comparison of benthic sampling procedures: Petersen vs. Makin corer. Water Research, 11, 597-601.

BARBOUR M.T., GERRITSEN G.E. SNYDER B.D. & STRIBLING J.B. (1999) Rapid bioassessment protocols for the use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish. 2nd edition. EPA 841-B-99-002. US Environmental Protection Agency, Office of Water, Washington, DC.

BASS J. A. B., WRIGHT J. F., CLARKE R. T., GUNN R. J. M. & DAVY-BOWKER J. (2000) Assessment of sampling methods for macroinvertebrates (RIVPACS) in deep watercourses. Environment Agency R&D Technical Report E134. 57pp.

BEETON A.M., CARR J.F & HILTUNEN J.K. (1965) Sampling efficiencies of three kinds of dredges in southern Lake Michigan. Proceedings of the 8th Conference on Great Lakes Research, 209.

BENJAMIN J. (1998) A comparative study of methods for sampling macroinvertebrates in Sussex Rifes. Unpublished report to Environment Agency, Southern Region. 103pp.

BLOCKSOM K.A. & FLOTEMERSCH J.E. (2005) Comparison of macroinvertebrate sampling methods for nonwadeable streams. Environmental Monitoring and Assessment, 102, 243–262.

BLOCKSOM K.A. & FLOTEMERSCH J.E. (2008) Field and laboratory performance characteristics of a new protocol for sampling riverine macroinvertebrate assemblages. River Research and Applications, 24, 373-387.

BRINKHURST R.O., CHUA K.E. & BATOOSINGH E. (1969) Modification in sampling procedures as applied to the studies on the bacteria and tubificid oligochaetes inhabiting aquatic sediments. Journal of the Fisheries Research Board of Canada, 26, 2581-2593.

BRETSCHKO G. & SCHÖNBAUER B. (1998) Quantitative sampling of the benthic fauna in a large, fast flowing river (Austrian Danube). Archiv für Hydrobiologie Supplement, 115, 195-211.

BURTON W. & FLANNAGAN J.F. (1973) An improved Ekman-type grab. Journal of the Fisheries Research Board of Canada, 30, 287-290.

CLARKE R.T., DAVY-BOWKER J., SANDIN L., FRIBERG N., JOHNSON R.K. & BIS B. (2006) Estimates and comparisons of the effects of sampling variation using 'national' macroinvertebrate sampling protocols on the precision of metrics used to assess ecological status. Hydrobiologia, 566, 477-503.

CLARKE R.T., FURSE M.T., GUNN R.J.M., WINDER J.M. & WRIGHT J.F. (2002) Sampling variation in macroinvertebrate data implications for river quality indices. Freshwater Biology, 47, 1735-1751.

DAVY-BOWKER J., JONES J.I. & MURPHY J.F. (2012) Standardisation of RIVPACS for deep rivers: Phase I – deriving a standard approach to deep river sampling. Environment Agency, Bristol.

DINES R.A. & MURRAY-BLIGH J.A.D. (2000) Quality Assurance and RIVPACS. In: Wright, J. F., Sutcliffe, D. W. & Furse, M. T. (Eds.), Assessing the biological quality of freshwaters; RIVPACS and other techniques. Freshwater Biological Association, Ambleside, 71-78.

DRAKE C.M. & ELLIOTT J.M. (1982) A comparative study of three air-lift samplers for sampling benthic macro-invertebrates in rivers. Freshwater Biology, 12, 511-533.

DRAKE C.M. & ELLIOTT J.M. (1983) A new quantitative air-lift sampler for collecting macroinvertebrates on stony bottoms in deep rivers. Freshwater Biology, 13, 545-559.

ELLIOTT J.M. & DRAKE C.M. (1981a) A comparative study of seven grabs used for sampling benthic macroinvertebrates in rivers. Freshwater Biology, 11, 99-120.

ELLIOTT J.M. & DRAKE C.M. (1981b) A comparative study of four dredges used for sampling benthic macroinvertebrates in rivers. Freshwater Biology, 11, 245-261.

ELLIOTT J.M., DRAKE C.M. & TULLETT P.A. (1980) The choice of a suitable sampler for benthic macroinvertebrates in deep rivers. Pollution Report to the Department of the Environment U.K. No. 8, 36-44.

ELLIOTT J.M., TULLETT P.A. & ELLIOT J.A. (1993) A new bibliography of samplers for freshwater benthic invertebrates. Ambleside, FBA Occasional Publication No 30. 92pp.

FAST A.W. (1968) A drag dredge. Progressive Fish Culturist, 30, 57-61.

FLOTEMERSCH J.E., BLOCKSOM K.A., HUTCHENS J.J. & ATUREY B.C. (2006) Development of a standardized large river bioassessment protocol (LR-BP) for macroinvertebrate assemblages. River Research and Applications, 22, 775-790.

FLOTEMERSCH J.E., STRIBLING J.B. & PAUL M.J. (2006) Concepts and Approaches for the Bioassessment of Non-wadeable Streams and Rivers. EPA 600-R-06-127. US Environmental Protection Agency, Cincinnati, Ohio.

FURSE M.T., CLARKE R.T., WINDER J.M., SYMES K.L., BLACKBURN J.H., GRIEVE, N.J. & GUNN R.J.M. (1995) Biological assessment methods: Controlling the quality of biological data. Package 1. The variability of data used for assessing the biological condition of rivers. R&D Note 412, Report to the National River Authority, Bristol, 139pp.

GAUFIN A.R., HARRIS E.K. & WALTER H.J. (1965) A statistical evaluation of stream bottom sampling data obtained from three standard samplers. Ecology, 37, 643-648.

HERRIG H. (1975) Der Bodensauger – ein neuartiges Gerät zur Entnahme von Sohlenproben aus großen Fließgewassern. (The slurp gun – a novel device for river bottom sampling. – in German). Deutsche Gewasserkdl. Mitteilungen19, 104-107.

HOLOPAINEN I.J. & SARVALA J. (1975) Efficiencies of two corers in sampling soft-bottom invertebrates. Annales Zoologica Fennica, 12, 280-284.

HOWMILLER R.P. (1971) A comparison of the effectiveness of Ekman and Ponar grabs. Transactions of the American Fisheries Society, 100, 560-564.

HUDSON P.L. (1970) Quantitative sampling with three benthic dredges. Transactions of the American Fisheries Society, 99, 603-607.

JACKSON M. J. (1997) *Sampling methods for studying macroinvertebrates in the littoral vegetation of shallow lakes*. Broads Authority, Norwich.

JONES J.I., BASS J.A.B. & DAVY-BOWKER J. (2005) *Review of methods for sampling invertebrates in deep rivers*. North South Shared Aquatic Resource (NS Share) Interim Report. 46pp

KAJAK Z. (1963) Analysis of quantitative benthic methods. Ekologia Polska (A), 11, 1-57.

KARLSSON M., BOHLIN T. & STETSON, J. (1976) Core sampling and flotation: two methods to reduce costs of a chironomid population study. Oikos, 27, 336-338.

KLEMM D. J., BLOCKSOM K. A., FULK F. A., HERLIHY A. T., HUGHES R. M., KAUFMANN P. R., PECK D. V., STODDARD J. L., THOENY W. T., GRIFFITH M. B. & DAVIS W. S. (2003) Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid-Atlantic Highlands streams. Environmental Management 31, 656-669.

LAZORCHAK J.M., HILL B.H., AVERILL D.K., PECK D.V. & KLEMM D.J. (2000) Environmental Monitoring and Assessment Program -Surface Waters: Field Operations and Methods for Measuring the Ecological Condition of Non-Wadeable Rivers and Streams. U.S. Environmental Protection Agency, Cincinnati OH

MACKEY A.P. (1972) An air-lift for sampling freshwater benthos. Oikos, 23, 413-415.

MCGARRIGLE M. L., LUCEY J. & CLABBY K. J. (1992) Biological assessment of river water quality in Ireland. In: River Water Quality – Ecological Assessment and Control. 371-393, Commission of the European Communities, EUR 14606 EN-FR, Luxembourg, 751pp.

MOULTON S.R.II, KENNEN J.G., GOLDSTEIN R.M. & HAMBROOK J.A. (2002) Revised protocols for sampling algal, invertebrate and fish communities as part of the national water-quality assessment program. USGS Open-File Report 02-150, US Geological Survey, Reston, Virginia.

MURRAY-BLIGH J. A. D., FURSE M. T., JONES F. H., GUNN R. J. M., DINES R. A. & WRIGHT J. F. (1997) Procedure for collecting and analysing macroinvertebrate samples for RIVPACS. Institute of Freshwater Ecology & Environment Agency, 155pp.

MILBRINK G. & WIEDERHOLM T. (1973) Sampling efficiency of four types of mud bottom samplers. Oikos, 24,479-482.

NEALE M.W., KNEEBONE N.T., BASS J.A.B., BLACKBURN J.H., CLARKE R.T., CORBIN T.A., DAVY-BOWKER J., GUNN R.J.M., FURSE M.T. & JONES J.I. (2006) Assessment of the Effectiveness and Suitability of Available Techniques for Sampling Invertebrates in Deep Rivers. North South Shared Aquatic Resource (NS Share) Final Report T1(A5.8) – 1.1. 97pp.

NORRIS R.H. (1980) An appraisal of an air-lift sampler for sampling stream macroinvertebrates. Bulletin of the Australian Society of Limnology, 7, 9-15.

OHIO EPA (1987) Biological Criteria for the Protection of Aquatic Life: Vol. III. Standardized Biological Field Sampling and Laboratory Methods for Assessing Fish and Macroinvertebrate Communities, Division of Water Quality Monitoring and Assessment, Ohio Environmental Protection Agency, Columbus, Ohio.

PEARSON R.G., LITTERICK M.R. & JONES N.V. (1973) An air-lift for quantitative sampling of the benthos. Freshwater Biology, 3, 309-315.

PEHOFER H. E. (1998) A new quantitative air-lift sampler for collecting invertebrates designed for operation in deep, fast-flowing gravelbed rivers. Archiv für Hydrobiologie Supplement, 101, 213-232.

PETERMEIER A. & SCHÖLL F. (1996) Das hyporheische Interstitial der Elbe – Methodenrecherche. Bundesanstalt für Gewässerkunde, Koblenz. BfG-1038.

PISCES CONSERVATION LTD. (1998) 'Species Diversity and Richness - Version 2' (1998) Software package: PISCES Conservation Ltd, Lymington, UK.

PREJS A. (1969) Differences in the abundance of benthos and reliability of its assessment in several lake habitats. Ekologia Polska (A), 17, 133-147.

PROBERT P.K. (1984) A comparison of macrofaunal samples taken by box corer and anchor-box dredge. NZOI Records, 4 (13), 149-157.

RAYSON M. (2000) Physiological Assessment of Deep Water Sampling Techniques by Biologists in the Environment Agency. Report to the Environment Agency by Optimal Performance Ltd, 19pp.

SLACK K.V., FERREIRA R.F. & AVERETT R.C. (1986) Comparison of four artificial substrates and the Ponar grab for benthic invertebrate collection. Water Resources Bulletin, 22, 237-248.

SLY P.G. (1969) Bottom sediment sampling. Proceedings of the 12th Conference on Great Lakes Research, 883-898.

STAŃICZYKOWSKA A. (1966) Some methodological problems in zoo-microbenthos studies. Ekologia Polska (A), 14, 382-393.

THAYER G.W., WILLIAMS R.B., PRICE T.J. & COLBY D.R. (1975) A large corer for quantitatively sampling benthos in shallow water. Limnology and Oceanography, 20, 474-480.

TURNER A.M. & TREXLER J.C. (1997) Sampling aquatic invertebrates from marshes: evaluating the options. Journal of the North American Benthological Society, 16, 694-709.

WASMUND E. (1932) Entwicklung und Verbesserung der Entnahmeapparatur für Bodenproben unter Wasser (Development and improvement of collecting apparatus for bottom sampling under water – in German). Archiv für Hydrobiologie, 2, 646-662.

WEBER C.I. (1973) Biological field and laboratory methods for measuring the quality of surface waters and effluents. Environmental Monitoring Series, U.S. Environmental Protection Agency. EPA 670/4-73001.

WRIGHT J. F., CLARKE R. T., GUNN R. J. M., BLACKBURN J. H. & DAVY-BOWKER J. (1999) Testing and further development of RIVPACS – Phase 3. Development of new RIVPACS methodologies . Stage 1. 138pp. Environment Agency.

List of abbreviations

ASPT	Average Biological Monitoring Working Party Score Per Taxon
BAMS	Biological Assessment Methods
BMWP	Biological Monitoring Working Party
EA	Environment Agency
EHS	Environment and Heritage Service of Northern Ireland
EPA	Environmental Protection Agency (of Republic of Ireland)
FBA	Freshwater Biological Association
GQA	General Quality Assessment
IFE	Institute of Freshwater Ecology
IRTU	Industrial Research and Technology Unit (Northern Ireland, UK)
LHPN	Long-handled pond net: a standard FBA pond net with a 1.5m long handle (referred to as a standard FBA long-handled pond net in Murray-Bligh *et al.* 1997), modified so that extensions can be fitted to increase the length to 4 m.
NS Share	North South Shared Aquatic Resource
NTAXA	Number of Biological Monitoring Working Party scoring Taxa
RICT	River Invertebrate Classification Tool
RIVPACS	River Invertebrate Prediction and Classification System
SEPA	Scottish Environmental Protection Agency
US EPA	United States Environment Protection Agency
USEPA-EMAP	United States Environmental Protection Agency, Environmental Monitoring and Assessment Program
USEPA-RBP	United States Environmental Protection Agency, Rapid Bioassessment Program
USGS-NAWQA	United States Geological Survey, National Water Quality Assessment Program

We are The Environment Agency. It's our job to look after your environment and make it **a better place** – for you, and for future generations.

Your environment is the air you breathe, the water you drink and the ground you walk on. Working with business, Government and society as a whole, we are making your environment cleaner and healthier.

The Environment Agency. Out there, making your environment a better place.