CHAPTER 3. DATA COLLECTION, STANDARDISATION AND STORAGE

Summary

This chapter discusses data collection and outlines the ecological reference condition protocol. It aims at standardising and improving the quality of data collected. The "environmental variables" incorporated into the reference condition field manual (Dallas 2000) are tabulated and their effects on invertebrate communities described. Technical considerations for sampling riverine macroinvertebrates using SASS are discussed, including aspects such as separate- versus combined-biotope sampling, single- versus multiple-season sampling, sampling frequency, habitat assessment as a tool for interpreting SASS Scores, taxonomic resolution, using SASS in non-perennial systems and the identification of environmental variables for future predictive systems. The data storage facility, namely the Rivers Database is briefly described.

3.1 INTRODUCTION

The quantity and quality of data collected within the RHP will ultimately determine the success and utility of the programme. During the initial stages of implementing such a programme, investment in terms of time and cost is generally high, but once the programme has a solid foundation, routine monitoring becomes quicker and simpler. Different provinces in South Africa are currently in different stages with respect to their involvement in the RHP. Mpumalanga, the province selected for the trial implementation of the RHP, is the most advanced. Others, e.g. Western Cape and KwaZulu Natal, have been doing biomonitoring for some time but not within a RHP framework, whilst others, e.g. Eastern Cape, have yet to begin regular biomonitoring programmes. There has, however, been great support and enthusiasm for the programme. The derivation of ecological reference conditions is a crucial interpretative tool within the RHP and hence every effort should be made within the provinces to ensure that data collected as part of the RHP may subsequently be utilised for deriving reference conditions and potentially used for predictive modelling in the future (See chapter 6).

3.2 STANDARDISATION OF DATA COLLECTION

The quality of data gathered requires that the biomonitoring protocol be standardised. In this way data collected from different regions and by different practitioners may be confidently compared and used. To standardise the collection of data for the RHP, an ecological reference condition field-manual has been written (Dallas 2000). It describes in detail the different components considered important for characterising a site, assessing the condition of a site, estimating various stream parameters, sampling invertebrates and water chemistry, and assessing habitat. The focus in this manual is the riverine macroinvertebrates and the characteristics of a site which need to be considered in relation to this component of the biota. Many of these components, however, also have relevance for other biotic components such as fish and riparian vegetation that are also part of the RHP.

In general terms, these site characteristics may be broadly referred to as "environmental variables". In developing the field-manual, an attempt was made to incorporate environmental variables which may:

- be important in characterising and describing a site,
- provide insight into the ecological condition or health of the site,
- potentially influence the distribution of riverine invertebrates, or
- provide greater interpretative power when undertaking invertebrate sampling using SASS.

The extent to which different environmental variables affect invertebrate distributions has not been tested and it is likely that their relative importance may vary from one geographic region to another. Ultimately, it would be useful to undertake analyses at a national level, incorporating each of the environmental variables such that a set which best characterise groups of reference sites can be identified and used for future predictive modelling. In this way the ecological condition in a river is assessed by predicting the invertebrate families expected in the absence of disturbance. It is likely however that the environmental variables used for model construction are a reflection of the spatial scale (e.g. national, regional, catchment or river) and hence extrapolation from one such scale to another is unlikely (Marchant *et al*. 1999) and predictive models may therefore need to be developed on a regional basis. The environmental variables in the field-manual (Dallas 2000) and their relevance with respect to ecological reference conditions and biomonitoring are listed in Table 3.1.

3.3 TECHNICAL CONSIDERATIONS FOR SAMPLING RIVERINE MACROINVERTEBRATES USING SASS

3.3.1 Separate- versus combined-biotope sampling

There has been much debate about the utility of sampling biotopes separately, with the obvious trade-off between time/cost and additional data gleaned. Ultimately the choice of separate- or combined-biotope sampling is dependent on the objectives of the biomonitoring or assessment programme. SASS-biotope availability may influence results of biomonitoring through preferential utilisation of different biotopes by invertebrates. SASS-defined biotopes include stones-in-current (SIC), stones-out-of-current (SOOC), marginal vegetation (MV), aquatic or instream vegetation (AQV), gravel (G), sand (S) and mud (M). Although there is limited information available on trends in biotope utilisation by invertebrates in South Africa, several studies have documented differences, particularly in taxon richness, amongst different biotopes (e.g. Pinder *et al.* 1987, Chessman *et al*. 1997, Dallas 1997, Kay *et al*. 1999). Translated into biotic indices such as SASS Scores, differences in number of taxa would affect the SASS4 Score, although it has been shown to have less effect on ASPT (Armitage *et al*. 1983, Dallas 1997). In AusRivAS, major biotopes are sampled separately because each biotope has a characteristic invertebrate community (Smith *et al*. 1999), and within a given region, differences among biotopes are greater than differences between sites. This is particularly important if predictive modelling is going to be developed as combined-biotope data adds considerable "noise" to the model and confounds interpretation of biological impairment (Parsons and Norris 1996).

Table 3.1 "Environmental variables" incorporated in the ecological reference condition field-manual (Dallas 2000) and their effects on invertebrate communities and therefore their relevance to ecological reference conditions and biomonitoring. The table is divided into three sections which relate directly to the sections in the field-manual.

	Component	**Detail**
SECTION A	Ecoregion	Ecoregions form level 1 of the hierarchical spatial framework aimed at identifying homogeneous regions. They are based on the classification of rivers into regions in which physical variables such as physiography, climate, geology, soils and potential natural vegetation are similar. It assumes that instream features such as the distribution of the biota or water chemistry are linked to these variables. Invertebrate communities differ amongst ecoregions. Since ecoregions require verification, the bioregion into which the site falls is also recorded.
	Bioregion	Bioregions are refined from biogeographic regions, which were based on broad historical distribution patterns of riverine macroinvertebrates, fish and riparian vegetation. Invertebrate communities differ amongst bioregions.
	Sub-region	This sub-regional classification reflects broad geomorphological characteristics and distribution patterns of components of the biota longitudinally down a river. Invertebrate communities differ amongst sub-regions.
	Catchment area	These all reflect the size of the river at the sampling site.
	Source distance	
	Stream order	
	Contour range/altitude	Altitudinal differences often result in water temperature differences and consequently in differences in invertebrate communities.
	Slope/gradient	Slope and gradient are correlated with stream flow. High-gradient streams are areas of rapid flow and erosion. Low-gradient streams have more sluggish flow and consequently are areas of deposition of finer sediments such as sand and silt. Invertebrate communities inhabiting high- versus low-gradient are often different.
	Geological type	Geology/lithology affects water chemistry, which in turn may affect invertebrate communities.
	Vegetation type	This refers to potential natural vegetation and it is a reflection of climate and soil. Certain types of vegetation leach humic substances into the water thereby altering the water chemistry. This may affect invertebrate communities.
	Hydrological type	Perennial, seasonal and ephemeral rivers are likely to have different invertebrate communities.
	Rainfall region	This may affect invertebrate communities and influences the choice of sampling season(s).

	Component	**Detail**
	Physical characteristics and geomorphology	This relates to shape of the valley, lateral mobility of the river, presence or absence of a macro-channel, channel pattern (single/multiple, braided, etc.) and channel type (bedrock/alluvial/mixed). The last two aspects, in particular, are likely to influence invertebrate community composition.
SECTION B	Catchment and land-use	Each of these components facilitates the assessment of the condition of the site by assessing the extent of anthropogenic activities at the site. The surrounding land-use and its potential water quality impacts, together with physical modifications to the channel, bed and bank, may negatively impact on the invertebrate community. Present status is a numerical method whereby these aspects are combined to get a general "status" of the site. Since reference conditions are based on least-impacted conditions, this section ensures that any potential impact is identified and its severity qualitatively ascertained.
	Water quality impacts	
	Channel condition	
	Present status	
SECTION C	Date visited	Invertebrate communities vary temporally, i.e. with time. Certain invertebrate taxa only occur in one or two seasons which could lead to misinterpretation if not accounted for when doing comparisons between monitoring and reference sites or condition. It is therefore necessary to understand the temporal effect of sampling period or season of sampling.
	Time of day	This may affect chemical variables such as pH, which varies over a 24 hour cycle.
	Water level	Any assessment undertaken during or shortly after a flood or spate event will not be a true reflection of the invertebrates present at the site.
	Rainfall?	
	Water turbidity	The extent to which the bottom of the river is visible affects one ability to estimate components such as substratum composition.
	Canopy cover	The extent of the canopy cover of the adjacent riparian vegetation affects water temperature and allochthonous (food produced outside the river) input of food. This in turn may affect invertebrate communities.
	Stream dimensions	The width of the active channel and surface water reflect the size of the river. The depths of the deep- and shallow-water biotopes may influence which invertebrates are present at a site.
	Substratum	The type of substratum (bedrock, boulder, cobble, pebble, gravel, sand and silt/mud) affects invertebrate communities.
	Biotopes	SASS biotopes which are present at a site affects the number of taxa recorded and hence the SASS4 Score. It is important to take this into account when assessing the degree of impairment at a site by comparing it with the reference condition. This should link to the SASS biotopes sampled.

	Component	Detail
	Invertebrate taxa	Invertebrates are sampled using the SASS method, either per biotope or by combining biotopes. Sampling taxa at the level of biotope allows greater interpretative power, in addition to increasing future predictive modelling capacity. Certain taxa are more commonly associated with certain SASS biotopes. By sampling biotopes separately, or at least in logical groups (e.g. SIC/SOOC, AQV/MV and G/S/M) it will be possible to interpret SASS scores relative to the number and type of biotope(s) sampled.
	Habitat assessment	By assessing habitat and invertebrates simultaneously, an attempt is made to differentiate between impacts resulting from water quality impairment and destruction of habitat. The most recent version (IHAS) attempts to incorporate differences in the availability and quality of SASS biotopes as an aid for the subsequent interpretation of SASS scores.
	Water chemistry	Invertebrates act as continuous monitors of the water they inhabit, enabling long-term analysis of both regular and intermittent discharges, variable concentrations of pollutants, single and multiple pollutants, and synergistic or antagonistic effects. Whilst it is impossible to measure the whole range of water chemistry variables, measuring a range of them will enable elevated or unnatural conditions to be ascertained. This information will also assist in the establishment of reference condition ranges for particular constituents which vary regionally such as pH, conductivity, turbidity etc.

In certain biomonitoring programs, sampling is confined to a particular biotope (Plafkin *et al.* 1989) in an attempt to reduce variability and predictive models based on samples from a single biotope were able to detect biological impairment (Parsons and Norris 1996). However, biased measurements may result because of among-biotope spatial variation in community composition. Similarly, if a human impact is specific to a particular biotope, and sampling is restricted to a different biotope, then the measurement of human impact may be biased (Kerans *et al.* 1992). For example, Pettigrove (1990, cited by Growns *et al.* 1997) noted that invertebrate communities in riffles and pools had differing sensitivities to different environmental factors, with nutrients and riparian vegetation having the greatest impact on riffle communities and turbidity having a greater impact on pool communities. Restricting analyses to one biotope also imposes limitations on the system in that it can only be used for sites where the selected biotope is present. For example many lowland rivers only have sand beds and marginal vegetation. Although there are clear limitations with using a "single biotope" sampling approach, it would be useful for to examine the potential of this approach in more detail.

With respect to SASS, it would be more advantageous to either sample biotopes separately or to develop a method whereby biotope availability can be factored into the interpretation of scores. This would need to be done on a regional basis because of intrinsic differences in the relative proportion and importance of biotopes between regions (e.g. marginal vegetation appears to be more important than stones-in-current in eastern

Cape rivers: pers. obs., M. Uys, Rhodes University).

In the reference condition manual, it has been suggested that, where available, five biotopes are sampled separately, as follows: stones-in-current (SIC), stones-out-of-current (SOOC), aquatic vegetation (AQV), marginal vegetation (MV) and gravel/sand/mud (G/S/M). Biomonitoring practitioners, however, have suggested that such biotope differentiation is impractical, and they commonly sample three biotope-groups separately as described below (C. Thirion, IWQS, DWAF, pers com.).

- Stones-in-current/stones-out-of-current (SIC/SOOC)
- Aquatic and marginal vegetation (AQV/MV)
- Gravel/sand/mud (G/S/M)

If time permits, SIC and SOOC could be sampled separately, and in rivers where AQV is considered a particularly important biotope, e.g. Western Cape, it is suggested that AQV and MV are sampled separately. It should be noted that, if biotopes are sampled separately, for those taxa which require the number of types to be recorded, namely baetids, hydropsychids and cased-caddis trichopterans, a site total should be recorded if different types were found in different biotopes. Sampling biotopes separately and then merging data from each biotope to obtain a SASS Score for the site often results in higher SASS Scores than if all biotopes were sampled together (Pers. Comm. C. Thirion, IWQS, DWAF). This is most likely a reflection of the greater amount of time spent recording the taxa, i.e. 15 minutes per biotope group (and hence a total of 45 minutes if three biotope-groups are sampled) versus 15 minute per site (Pers. Comm. C. Dickens, Umgeni Water). The influence of biotope availability is examined in section 4.4.3.

3.3.2 Single- versus multiple-season sampling and sampling frequency

When invertebrate communities are used for biomonitoring, temporal variation may influence judgement as to whether or not a site is disturbed. For example, Linke *et al*. (1999) noted consistent differences in number of taxa and the biotic index between summer and winter samples. Biomonitoring based on these results would indicate better water quality in the same streams in winter relative to summer even though the degree of disturbance had remained constant. Two questions related to seasonality arise:

- How many seasons' data should a reference condition be based on?
- How can monitoring data be compared to reference condition data such that season of sampling does not affect data interpretation?

Reference condition: Any biomonitoring programme that attempts to standardise sampling protocols must attempt to ascertain the importance of seasonality. A minimum number of samples must therefore be taken through time to allow collection of adequate invertebrate information for "reference" site classification or derivation of the ecological condition of a group of similar reference sites. Based on experience in the United Kingdom and Australia, it is recommended that at least two, and preferably three, seasons are sampled each year, excluding the highest rainfall period. In South Africa, in summer-rainfall regions this would then be autumn, winter and spring, and in winter-rainfall regions, spring, summer and autumn. The classification of reference sites is then undertaken on the seasonally-combined site taxa data. Seasonal sampling of reference sites should continue until an understanding is gained on the influence of season on the interpretation of biomonitoring data. At such a stage, the frequency of sampling of reference sites may be decreased to once per year, in the season deemed most appropriate through data analyses.

Comparing monitoring data to reference condition data: It has been shown that predictive models based on a single season have greater predictive power than those based on combined-season data (Smith *et al*. 1999, Linke *et al*. 1999). Since such predictive models are not yet available for South African rivers, an alternative method, whereby sampling season is taken into account, needs to be developed. Both the variability in SASS Scores and seasonal differences in the presence of taxa need to be taken into account when interpreting data. The influence of seasonality is examined in section 4.4.4

3.3.3 Habitat assessment as a tool for interpreting SASS Scores

In the broadest sense, "habitat" incorporates all aspects of physical and chemical constituents along with the biotic interactions. In relation to reference conditions for invertebrates and to SASS assessments, the definition of "habitat" is narrowed to the quality of the instream habitat that influences the structure and function of the riverine invertebrate community.

In the past, two habitat assessment methods have been used, namely the Habitat Assessment Matrix (HAM, Roux 1993) and Habitat Quality Index (HQI, Thirion 1995). Both of these evaluate the physical habitat of the site and include information on bottom substrate/available cover, embeddedness, biotope diversity categories, velocity/depth categories, area of bottom affected by scouring and deposition, pool/riffle and run/bend ratios, bank erosion potential, bank vegetation stability and streamside cover (dominant vegetation).

It is considered important to assess the habitat available to invertebrates since this is likely to affect SASS Scores. The terms habitat and biotope are often used interchangeably and in relation to SASS, either one relates to the potential areas that are available for habitation by invertebrates. Since biotopes have more relevance for invertebrates, this term is used in preference to "habitat". The quality and quantity of available biotopes affects the structure and composition of invertebrate communities and the number of taxa present at a site appears to be related to the number of biotopes present (Dallas 1997). In terms of SASS Scores, biotope availability influences SASS4 Scores, although the effect on ASPT is less pronounced (Dallas 1997).

When one or more biotopes does not exist or is inaccessible, particularly those likely to support a substantial number of invertebrate taxa, such as stones-in-current and marginal vegetation, and if the absence of these biotopes is not accounted for, misinterpretation of SASS scores is likely to occur. In interpreting SASS results one needs to be confident that the effect observed is the result of impaired water quality, and hence river health, as opposed to limited habitats or biotopes. The extreme case would be if SASS were used to assess the water quality in canalised rivers. Such systems have virtually no biotopes present for habitation by invertebrates, and even if the water was of excellent quality, SASS scores, particularly the SASS4 Score, would be undervalued.

Final conclusions regarding the presence and degree of biological impairment should thus include an evaluation or factoring in of biotope quantity and quality to determine the extent to which biotope availability may be a limiting factor. A system currently under development, the Invertebrate Habitat Assessment System (IHAS, McMillan 1998) attempts to incorporate the availability and quality of biotopes so that it may be used as a tool to assist with the interpretation of SASS data. This method has yet to be tested and its usefulness in assisting with SASS data interpretation evaluated. The utility of the IHAS technique is examined in section 4.4.3.4.

3.3.4 Taxonomic resolution

SASS is based largely on the taxonomic level of family, with some taxa, e.g. Oligochaete at a higher level, and others, e.g. Baetidae and Hydropsychidae, incorporated in a sliding-scale manner based on the number of types. The number of types of cased-caddis trichopterans is also recorded. On the basis of Marchant *et al.*s' (1995) findings which showed that the family level identification provided an adequate taxonomic discrimination from which to group sites based on their invertebrate communities, the AusRivAs programme uses family level taxonomic data. Whilst no research has been undertaken in South Africa comparing family-level SASS data with studies of greater taxonomic resolution, it is likely that SASS would perform satisfactorily. SASS has been shown to adequately differentiate between sites of different water quality in a way comparable to a protocol involving intensive box-sampling and laboratory sorting (Dallas & Day, in prep.).

Incorporating a lower taxonomic level such as genus or species in a field-based rapid bioassessment method is not practical. For purposes of biomonitoring, and the derivation of reference conditions as an interpretative tool for such biomonitoring, family-level taxonomy is adequate. Unlike studies focusing on biodiversity and conservation, where rare taxa are important, biomonitoring aims to establish the ecological condition or health of a river, using invertebrates as indicators. Such rare taxa are not important for biomonitoring and increasing taxonomic resolution may in fact increase the noise by increasing the variability. Variability in both the biotic and physicochemical components have always been an important consideration in biomonitoring and environmental variability is a fundamental problem in assessing changes in water quality through time and space (Norris & Georges 1993).

3.3.5 Using SASS in non-perennial systems

SASS was developed for use in running waters. In such systems it has proved very useful in assessing water quality and general river health. It is tempting to use SASS in systems other than running-water systems, such as seasonal or ephemeral rivers, where isolated pools may be sampled, or in wetland areas. In both these cases, it is likely that SASS results would be erroneous and open to misinterpretation, and subsequent conclusions drawn with respect to water quality and river health misguided. Until research is conducted on the use of SASS, or a modified SASS method, for use in such systems, SASS assessments should be confined to perennial waters.