1.3 APPROACHES FOR DERIVING ECOLOGICAL REFERENCE CONDITIONS FOR RIVERINE MACROINVERTEBRATES

A reference condition is normally derived from several reference sites, which are within a homogenous region or faunal grouping. There are essentially two approaches, namely multimetric and multivariate, for defining reference conditions with which test or monitoring sites can be compared. The two approaches are fundamentally different even though they begin with the same premise and require the same data (Reynoldson *et al*. 1997).

1.3.1 Multimetric approaches

The multimetric approach classifies reference sites based on geographic and physical attributes. Geographic regions, termed ecoregions, are predefined largely using geomorphological characteristics such as climate, physiography, geology, soils and vegetation (Omernik 1987, cited by Reynoldson *et al*. 1997). This approach assumes that the test site characteristics match the chosen ecoregion reference sites (Reynoldson *et al*. 1997). Naturally occurring biotic assemblages, as components of the ecosystem, would be *expected* to differ among ecoregions but to be relatively similar within a given ecoregion. The ecoregion concept thus provides a geographic framework for efficient management of aquatic ecosystems and their components.

Within an ecoregion, additional qualifiers such as stream size, hydrologic regime, elevation, and natural riparian vegetation need to be considered for partitioning variability (Barbour *et al*. 1999). The aim is to establish homogeneous regions within which biomonitoring is conducted and for which ecological reference conditions are derived. Details of the spatial framework established for South Africa are provided in section 1.3. Metrics such as measures of richness (e.g. total number of taxa), composition [e.g. Ephemeroptera: Plecoptera: Trichoptera (EPT) ratio], tolerance/intolerance (e.g. % tolerant taxa), feeding (e.g. % Filterers etc.] and habit (e.g. % clingers) or indices [e.g. SASS4 Scores, Average Score Per Taxon (ASPT)] are then interpreted within the homogeneous regions. Several metrics are tabulated (Table 1.1) below to illustrate the range of possible metrics.

One of the advantages of metrics and indices is that they formalise what any good biologist, familiar with local biota, knows about the biological condition of a stream (Fore *et al*. 1996). These narrative and numeric indexes communicate biological condition to policy makers and concerned citizens and thus provide a scientific basis for the management decisions that affect aquatic resources (Fore *et al*. 1996).

1.3.2 Multivariate approaches

Multivariate approaches classify reference sites using multivariate analysis of macroinvertebrate fauna (Reynoldson *et al*. 1997). They make no *a priori* assumptions about the similarity of macroinvertebrate communities at different sites based on either physical or chemical descriptions. Rather, faunal data are used to group sites that have similar taxonomic composition, thus providing an objective way of grouping reference sites with similar macroinvertebrate communities. The multivariate approach does not assume that test sites exactly match reference site groups, but instead calculates the probability of belonging to each of the

Table 1.1 Definitions of potential benthic metrics and predicted direction of metric response to increasing perturbation (from Barbour *et al*. 1999).

Category	**Metric**	**Definition**	**Predicted response to increasing disturbance**
Richness measures	Total number of taxa	Measures the overall variety of the macroinvertebrate assemblage	Decrease
	Number of Ephemeroptera/Plecoptera/ Trichoptera (EPT) taxa	Number of taxa in the insect orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies)	Decrease
	Number of Ephemeroptera taxa	Number of mayfly taxa (usually genus or species level)	Decrease
	Number of Plecoptera taxa	Number of stonefly taxa (usually genus of species level)	Decrease
	Number of Trichoptera taxa	Number of caddisfly taxa (usually genus or species level)	Decrease
Composition measures	% EPT	Percent of the composite of mayfly, stonefly, and caddisfly larvae	Decrease
	% Ephemeroptera	Percent of mayfly nymphs	Decrease
Tolerance/ Intolerance measures	Number of Intolerant taxa	Taxa richness of those organisms considered to be sensitive to perturbation	Decrease
	% Tolerant Organisms	Percent of macrobenthos considered to be tolerant of various types of perturbation	Increase
	% Dominant Taxon	Measures the dominance of the single most abundant taxon. Can be calculated as dominant 2, 3, 4, or 5 taxa.	Increase
Feeding measures	% Filterers	Percent of the macrobenthos that filter FPOM from either the water column or sediment	Variable
	% Grazers and Scrapers	Percent of the macrobenthos that scrape or graze upon periphyton	Decrease
Habit measures	Number of Clinger Taxa	Number of taxa of insects	Decrease
	% Clingers	Percent of insects having fixed retreats or adaptations for attachment to surfaces in flowing water.	Decrease

groups (Reynoldson *et al*. 1997).

Both the United Kingdom (Wright *et al.* 1993) and Australia (Smith *et al.* 1999) have adopted the multivariate approach within their biomonitoring programmes, respectively RIVPACS (River Invertebrate Prediction and Classification System) and AusRivAS (Australian River Assessment System). They are described below.

1.3.2.1 RIVPACS

The RIVPACS approach has been summarised by Eekhout *et al*. (1996) as follows:

- "least impacted" sites are selected on the basis of water quality and expert opinion.
- Water chemistry and macroinvertebrate data are collected from the major habitats (combined in a single sample) at each site during spring, summer and autumn. The macroinvertebrate communities present at each site are sampled using a rapid assessment procedure, usually the British Monitoring Working Party (BMWP) scoring system.
- Macroinvertebrates are identified to the taxonomic level of species.
- Multivariate analysis of macroinvertebrate data is undertaken to establish groupings of sites with similar macroinvertebrate communities.
- The environmental characteristics distinguishing each group of reference sites are identified.
- Macroinvertebrate communities at monitoring sites are compared with those at reference sites with the most similar environmental characteristics.

This results in the generation of Observed/Expected ratios (O/E) which enable the degree of degradation of the test or monitoring site to be established. Additional features of RIVPACS include the option to run the analyses for one, two or three seasons, at species- or family-level, using presence/absence or $\log_{10}$ categories of abundance data and with the biological index, BMWP. The refinement of RIVPACS is an on-going process and future challenges include the development of early warning indices, linking RIVPACS and habitat components and the development of indices to detect particular types of environmental stress (Wright 2000).

1.3.2.2 AusRivAS

AusRivAS models were developed, using macroinvertebrates as indicators, to assess the ecological condition of rivers in Western Australia (Smith *et al*. 1999). AusRivAS is based on the RIVPACS system but with two modifications. Macroinvertebrates are only identified to family level, an approach supported by the findings of Marchant *et al*. (1995) who showed that the family level identification provides an adequate taxonomic discrimination from which to group sites based on their macroinvertebrate communities. Second, major habitats (channel, riffle, etc.) are sampled and processed separately.

As with RIVPACS, AusRivAS models assess the ecological condition in a river by predicting the macroinvertebrate families expected in the absence of disturbance (Smith *et al*. 1999). Predictions are derived from a set of environmental measurements used to characterise the site. A predicted macroinvertebrate assemblage is compared with the actual assemblage and the ratio of observed/expected (*O/E*) families is used as a measure of ecological condition (Wright *et al*. 1993, Parsons & Norris 1996). Whilst outside the scope of the reference condition project, it was considered useful to summarise the process followed in the construction of the AusRivAs model. Details are provided on the process and statistical techniques used in each step.

1.3.2.3 Model construction in AusRivAS

This section has been summarised from Smith *et al*. (1999). Wet-, dry- and combined-season models were produced for each habitat. Data collected during the first two sampling events were used to create the models, while data from the last two sampling events were used for validation. Model building occurred in five steps:

1. Reference sites were classified into groups with similar macroinvertebrate communities. An agglomerative hierarchical fusion technique (UPGMA, Unweighted Pair-Group aritMetic Averaging) from PATN analysis package was used with the Bray Curtis association measure and presence/absence data.. TWINSPAN was used if adequate discrimination of sites was not achieved. Sites with fewer than six families were excluded and families occurring at <5% of the sites were also excluded, so that classifications were not affected by rare families. Resultant groups should not contain less than five sites as recommended by Wright *et al*. (1993).

2. Once the optimal classification was chosen, stepwise Multiple Discriminant Function Analysis (DFA) was used to identify which environmental variables best discriminated between groups in the classification. This was done using the STEPDISC procedure in the SAS statistical package, with all environmental variables being checked for normality before use in DFA. Environmental variables such as altitude, distance from source, and latitude/longitude often make good predictor variables. Variables commonly affected by anthropogenic activity, such as turbidity, dissolved oxygen and phosphorus concentration, were omitted. Variables which contributed significantly ($P<0.05$) to group discrimination were classed as predictor values. Alternatively, the relationship between environmental and macroinvertebrate data can be determined by performing ordination followed by Principle Axis Correlation. Semi-strong hybrid multidimensional scaling (SSH) is performed on macroinvertebrate data with dimensions in ordination space added until an acceptable stress level (< 0.2) is achieved.

3. The DISCRIM procedure in SAS was used to incorporate predictor variables into a discriminant function and to assign sites to groups identified in the classification. The CROSSVALIDAT option was used to check whether sites were assigned to the correct group. Misallocations were examined and, if they were outliers or atypical, they were dropped from the classification.

4. The probability of each family occurring at a site was calculated by multiplying the probability of a site belonging to a classification group by the probability of a family occurring in that group and then summing the products to give the number of families expected (*E*). Only families with a probability >0.5 were used in calculations of *E* because uncommon families contribute very little to *E*-values and their

occurrence causes substantial stochastic fluctuations in *O/E* ratios. The frequency of occurrence of each taxon in each of the reference site classification groups is then calculated. Multiplying a taxon's frequency of occurrence in a classification group by the probability of a test site belonging to that group and summing the results for all of the groups in the classification, gives the probability of a taxon occurring at that test site.

5. Using a preliminary model *O/E* ratios of reference sites were calculated. Sites with *O/E* ratios <0.75 were removed from the classification on the assumption that they were not true reference sites and steps two to four were repeated for the final model.

AusRivAS also calculates the expected SIGNAL score for a site. SIGNAL (Chessman 1995) is a system that assigns a grade to each macroinvertebrate family based on its sensitivity to pollution and is similar to the SASS4 used in South Africa. The sum of all the grades at a site are then divided by the number of families to give an average SIGNAL score for the site. For each site the probabilities of occurrence of each taxon are multiplied by the SIGNAL grade for each taxon. The expected SIGNAL score for a site is then calculated by summing the weighted SIGNAL grades and dividing the total by the number of expected families. A banding system has also been developed to interpret the *O/E* scores.

1.3.2.4 The South African approach

The approach adopted in South Africa is a regional one, whereby a hierarchical spatial classification scheme sub-divides the country in a logical and ecologically-meaningful way so that variation between rivers in the country is best accounted for (Eekhout *et al*. 1996). As such we conform to the multimetric approach with sites classified *a priori* into homogeneous regions. The regional classification serves as a framework to aid the selection of both reference and monitoring sites, and hence the derivation of ecological reference conditions. Eekhout *et al*. (1996) stressed that the adoption of a regional approach does not preclude the eventual transference to a multivariate approach, whereas the opposite is true. The two approaches with respect to riverine macroinvertebrates in rivers in Mpumalanga have been examined in some detail in chapter 4.

1.4 A SPATIAL FRAMEWORK FOR DELINEATING HOMOGENEOUS REGIONS

1.4.1 Background

South Africa has a varied climate (and hence hydrological type), geology (and hence water chemistry) and geomorphology (and hence channel type, substratum composition, erosion potential). Variation in these factors, both among and within rivers, together with natural biogeographic differences in the distribution of riverine biota, may lead to biotic differences. Such differences need to be taken into account when implementing a national biomonitoring programme and deriving ecological reference conditions. In order to take account of these regional differences, the approach adopted in South Africa, was to develop a three-tiered hierarchical spatial framework (Figure 1.1).

1.4.2 A hierarchical spatial framework

The spatial framework was developed during a workshop in January 1996 (Brown *et al*. 1996) and it has been applied within the ecological reference condition project (details are given in Dallas & Fowler 2000). During this workshop, the biogeographic regions of Eekhout *et al*. (1997) were modified into bioregions. Sub-regions, which were largely a reflection of the geomorphological nature or zonation of rivers, and which represent level 2 of the hierarchy, were also delineated. As the bioregions and sub-regions were derived on the basis of limited data and professional judgement, they required verification. Level 3, namely river types, which related to aspects such as river size, hydrological type, geomorphological, chemical and biological characteristics, were not identified at the workshop, but were recognised as being important. This spatial framework is discussed in more detail below and subsequent developments related to each aspect are outlined.

1.4.2.1 Level 1: Ecoregions or bioregions

Bioregions (Brown *et al*. 1996) were refined from biogeographic regions, which were based on the broad historical distribution patterns of riverine macroinvertebrates, fish and riparian vegetation (Eekhout *et al*. 1997). The modifications to the biogeographic regions were made by specialists with a knowledge of macroinvertebrate, fish and/or riparian vegetation within each region. Subsequent to this, and as part of the development of procedures for the determination of the Ecological Reserve Project, an ecological classification or river typing method (Kleynhans *et al.* 1998) was followed which was a "top-down" approach and was based on a technique adopted in the United States (Omernik 1987, 1995; Omernik and Bailey 1997). In this top-down approach physical variables such as physiography, climate, geology, soils and potential natural vegetation are used to classify rivers into ecoregions. The assumption in an "ecoregion approach" is that instream features such as the distribution of the biota or water chemistry are intimately linked to these variables in the order in which they were placed in the classification hierarchy (Eekhout *et al.* 1997).

The differences between bioregions and ecoregions have not been established, and it is likely that one or the other will be more applicable to certain regions with respect to biomonitoring and the RHP. Ecoregions have been adopted for this study, which is based in Mpumalanga, although the bioregional classification may be more appropriate for other regions (for example the Western Cape, pers. comm., D. Van Niewenhuizen, Freshwater Research Unit, University of Cape Town). The ecological significance of ecoregions will need to be verified as data are collected within the RHP and it will most likely be an iterative process, with ecoregions or bioregions becoming more refined over time. Further ecoregional sub-levels within ecoregion Level I (Ecoregions Levels II and II) are envisaged for the development of procedures for the determination of the ecological reserve and these may also be incorporated into the biomonitoring framework if they provide additional clarity with regard to biota and reference conditions.

1.4.2.2 Level 2: Sub-regions or geomorphological zones

This level 2, sub-regional classification reflects broad geomorphological characteristics and distribution patterns of components of the biota. Rivers are longitudinally divided into the following zones: Source zone, Mountain headwater stream, Mountain stream, Foothills (cobble bed), Foothills (gravel bed) (previously termed Transitional) and Lowland sand bed or Lowland floodplain (Wadeson 1999). Three other geomorphological zones associated with a rejuvenated profile, namely Upland Flood Plain, Rejuvenated bedrock fall/cascade and Rejuvenated foothills, were also proposed. The characteristics of each geomorphological zone in terms of gradient and diagnostic channel features have been tabulated in Dallas & Fowler (2000).

1.4.2.3 Level 3: River types

Level 3 in the hierarchy is the identification of river types, which is performed using factors such as river size (e.g. stream width, stream order etc.), hydrological type (ephemeral, seasonal or perennial), geomorphological characteristics (channel pattern, substratum composition) and other chemical and biological factors. Identification of river types within each sub-region is a time-consuming process and it is envisaged that this could be done gradually by local experts as the RHP is implemented within different parts of the country (Eekhout *et al*. 1996).

1.5 STRUCTURE OF THIS REPORT

The process of deriving ecological reference conditions for riverine macroinvertebrates has been followed for rivers in Mpumalanga and all analyses performed in subsequent chapters are on Mpumalanga data. The aim of this report is to expand on issues of relevance to reference conditions, to describe problems experienced during the process of deriving reference conditions, and to suggest ways in which the protocol can be applied in other geographic regions. This report attempts to:

1. Describe the selection of potential reference sites and the ground-truthing or preliminary site screening process (Chapter 2).
2. Outline the type of data which need to be collected, with emphasis on standardising data collection such that the quality and quantity of this data is suitable for future advances of the RHP. It outlines technical considerations for sampling invertebrates using SASS, including seasonal aspects, separate- versus combined-biotope sampling, habitat assessment, taxonomic resolution, using SASS in non-perennial systems and identification of environmental variables for future predictive modelling. The storage of data within the Rivers Database is discussed (Chapter 3).
3. Provide details on site characterisation and data analysis. The procedure for analysing both the invertebrate and environmental data are described, together with a description of the statistical methods used. The influence of biotope availability and season of sampling is examined and procedures for factoring them in explored (Chapter 4).
4. Derive ecological reference conditions for riverine macroinvertebrates for selected river types within Mpumalanga and to characterise each of these river types with respect to abiotic factors (Chapter 5).
5. Provide recommendations and discuss aspects related to biomonitoring which will enhance the utility of the RHP and the use of ecological reference conditions within the RHP (Chapter 6). A schematic diagram of the suggested protocol for deriving reference conditions as examined in this report is provided.

Level 1
Ecoregions or bioregions

- Based on abiotic physical variables such as physiography, climate, geology, soils and potential natural vegetation (ecoregions) or biota (bioregions).
- Possible ecoregion sub-levels (II and III).

▼

Level 2
Sub-regions or geomorphological zones

- reflects broad geomorphological characteristics and distribution patterns of biotic components. Geomorphological zones include:
  - Source zone
  - Mountain headwater stream
  - Mountain stream
  - Foothills (cobble bed)
  - Foothills (gravel bed)
  - Lowland sand be or Lowland floodplain
  - Upland Flood Plain
  - Rejuvenated bedrock fall/cascade
  - Rejuvenated foothills.

▼

Level 3
River types

- Identified using factors such as:
  - river size (stream width, stream order, distance from source)
  - hydrological type (ephemeral, seasonal or perennial)
  - geomorphological characteristics (channel pattern, substratum composition)
  - other chemical and biological factors.
- Gradual process with river types being identified within subregions as the RHP is implemented within each geographical region.

Figure 1.1 A three-tiered hierarchical spatial framework indicating the components incorporated at each level and, in the case of sub-regions, the different sub-regions.