THE APPLICATION OF METHODS OF UNCERTAIN REASONING TO THE BIOLOGICAL CLASSIFICATION OF RIVER WATER QUALITY

MICHAEL BOYD

June 1995

THE APPLICATION OF METHODS OF UNCERTAIN REASONING TO THE BIOLOGICAL CLASSIFICATION OF RIVER WATER QUALITY

MICHAEL BOYD

Doctor of Philosophy

THE UNIVERSITY OF ASTON IN BIRMINGHAM

June 1995

The University of Aston in Birmingham

The Application of Methods of Uncertain Reasoning to the Biological Classification of River Water Quality

Michael Boyd

Doctor of Philosophy
1995

SUMMARY

This thesis presents an investigation into the application of methods of uncertain reasoning to the biological classification of river water quality.

Existing biological methods for reporting river water quality are critically evaluated, and the adoption of a discrete biological classification scheme advocated. Reasoning methods for managing uncertainty are explained, in which the Bayesian and Dempster-Shafer calculi are cited as primary numerical schemes.

Elicitation of qualitative knowledge on benthic invertebrates is described. The specificity of benthic response to changes in water quality leads to the adoption of a sensor model of data interpretation, in which a reference set of taxa provide probabilistic support for the biological classes. The significance of sensor states, including that of absence, is shown. Novel techniques of directly eliciting the required uncertainty measures are presented.

Bayesian and Dempster-Shafer calculi were used to combine the evidence provided by the sensors. The performance of these automatic classifiers was compared with the expert's own discrete classification of sampled sites. Variations of sensor data weighting, combination order and belief representation were examined for their effect on classification performance. The behaviour of the calculi under evidential conflict and alternative combination rules was investigated.

Small variations in evidential weight and the inclusion of evidence from sensors absent from a sample improved classification performance for Bayesian belief and support for singleton hypotheses. For simple support, inclusion of absent evidence decreased classification rate. The performance of Dempster-Shafer classification using consonant belief functions was comparable to Bayesian and singleton belief.

Recommendations are made for further work in biological classification using uncertain reasoning methods, including the combination of multiple-expert opinion, the use of Bayesian networks, and the integration of classification software within a decision support system for water quality assessment.

Keywords: Bayes, Dempster-Shafer Reasoning, Knowledge Elicitation, Biological Surveillance, River Water Quality.

Acknowledgements

I would like to acknowledge several people for their help at various times throughout this project. Sincere thanks goes to Bill Walley for originating the investigation into the use of artificial intelligence techniques in the domain of biological surveillance, and providing the impetus for this study. Bill has been a source of ideas, encouragement, guidance and critical appraisal, particularly in his scrutiny of this thesis at various stages of its development.

Sincere thanks to Bert Hawkes, who as domain expert gave his time and expertise freely throughout the knowledge elicitation exercises and indeed the whole project, and for his examination and classification of many benthic samples. It has been a privilege to be associated with someone of his academic stature. His kindness and encouragement have helped me to persist with the writing of this thesis.

Also I would like to express particular gratitude to John Elgy, for his help and continued interest in my submission. John's incisive and constructive criticism of the thesis, alongside those of Dr. John Murray-Bligh, has been of great value. My current supervisor, David Just was also extremely supportive of my efforts to submit this work.

I am grateful to Viki Hirst and her colleagues at the Environment Agency (North East Region), the eventual successors of the Yorkshire Water Authority, for allowing me access to the original records of the invertebrate surveys referred to in this thesis.

I would also like to express my appreciation of former colleagues in Civil Engineering for their companionship, ideas, and support during my time at Aston: particularly Brendan, Rob, Dennis, Peter Hedges for general encouragement and advice on structuring the thesis, Dr. Roger Kettle, former head of the Department of Civil Engineering, and my current colleagues at the University of Central England, particularly Ian, Stan, David, and Stephen.

Thanks also to my parents John and Agnes Boyd, sisters Marie and Bernadette and my brother John, and to Olive and Victor. Finally, and most important, my deepest thanks goes to my wife Christine, for her constant love, support, encouragement and patience throughout this lengthy submission, and to our two wonderful daughters Rachel and Hannah.

Contents

... **Contents**

... **Contents**

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Background

The biological assessment of river water quality is an activity requiring skill and expertise, relying on the knowledge and experience of practitioners in the field. In practice, biological survey data must be reduced via indices or score systems to summary form suitable for use by managers or decision makers. Some of these procedures use statistical techniques to give quality measures that are often viewed as objective. Others, such as biotic indices, draw on theoretical and experiential knowledge of the ecology of aquatic organisms and their responses to changes in water quality. For this latter approach the available knowledge is usually incomplete, relationships are inexact, and the conclusions should therefore reflect the inherent uncertainty. So far, this has been ignored for biotic indices, which historically have been designed for ease of use and arithmetic simplicity.

Artificial intelligence is an area of study concerned with using computational methods for solving problems that apparently require human intelligence. Probabilistic and related uncertain reasoning methods are a branch of artificial intelligence that provide mathematically coherent procedures by which imprecise and uncertain knowledge can be integrated and used to make rational and 'fair' decisions in some narrow domain.

This project and thesis encompass a detailed investigation of the application of these techniques for the biological classification of river water quality. It is a multifaceted and exploratory study into problems of defining and classifying water quality, the modelling and representation of benthic sensor data as probabilistic knowledge, the elicitation of this knowledge from an expert in the field, and the behaviour and performance of numerical decision algorithms in this domain.

1.2 Role of Uncertain Reasoning

Human experts can reason and make decisions when faced with uncertain facts, an ability that partly underpins their expertise. According to Hart (1989): "much of the skill in judgement lies in weighing up the relative merits of data, facts, guesses and hypotheses, etc. and using a plausible line of reasoning with them". In assessing biological river water quality, the prime source of data is the biotic communities themselves. The problem to be solved, following a plausible line of reasoning, is to determine the biological quality of the river water.

The reasoning systems developed in this study attempt to emulate the expert's ability to decide the quality of river water from samples of benthic invertebrate communities. In this regard, they too must reason under uncertainty. Incompleteness in benthic data arises from errors in the sampling process, integrity of the sample data, effects of seasonality or local conditions on occurrence and distribution of the taxa, and the sensitivity to varying levels of river water quality. Natural variations exist in the susceptibility to pollutants of individual animals and between different populations. Gaps in knowledge of deeper causal mechanisms (for example sensitivities to different pollutants, or predator-prey relationships) affecting most benthic organisms will also contribute to uncertainty. Aquatic communities are not part of a predictable, deterministic environment: rather they are subject to stochastic events (Jeffries and Mills, 1990).

The systems described in this dissertation improve on the traditional use of biotic indices and score systems in several ways. First, they are mathematically coherent procedures for manipulating uncertain information, which allow rational decisions to be made regarding river water quality. Secondly, they encode ecological knowledge directly elicited from an expert in the field of biological surveillance. This knowledge models the occurrence of taxa across a range of water qualities and the importance of different levels of abundance, including the use of information provided by the absence of taxa. Thirdly, water quality is reported in terms of a discrete classification, mirroring the NWC chemical classes. It is argued in this thesis that discrete systems are more effective for communicating information on water quality.

1.3 Research Objectives

The major objectives of this research project were to:

(i) review methods of biological surveillance and critically appraise classification schemes used for reporting river water quality

(ii) construct, via a process of knowledge acquisition and consultation with an acknowledged expert in the field of biological surveillance, a model of the interpretation of benthic data in terms of a biological classification scheme

(iii) investigate the application of uncertain reasoning calculi to emulate the expert's ability to classify river water quality by interpreting benthic data

(iv) develop computer programs that incorporate decision algorithms for automating biological classification, which could be used as part of an overall decision support system for biological surveillance of river water quality

(v) investigate by a programme of computational experiments the performance of the various calculi in matching the expert's classification.

1.4 The BERT System

Aston University has through its former Applied Hydrobiology Section (Department of Biological Sciences) a wide experience of teaching and research into the biological surveillance of rivers. Bert Hawkes, for many a years a Reader within this section and now with the Department of Civil Engineering, has contributed numerous scholarly papers and articles on the subject. His expertise in this field is widely acknowledged. The originator of this project, Bill Walley, for many years a senior tutor in Civil Engineering at Aston University, foresaw the possibilities of applying artificial intelligence techniques to this domain. To this end he initiated the setting up of a studentship for research into this area.

The work described in this thesis is one part of a more ambitious project to design and develop a knowledge-based system for assessing river pollution, in which biological methods would play a major role. The aims of this system, called BERT (Benthic Ecology Response Translator), were to identify (i) the spatial and temporal changes of pollution in rivers, (ii) the types and likely concentrations of pollutants and (iii) the most likely sources of pollutants (Walley *et al.*, 1992b).

Originally, the work done as part of this studentship was directed towards the development of this overall system. One possible design for the system developed by the author is briefly described in Chapter 8. However, the BERT system as originally conceived has yet to be realised, since the direction of the project changed from the development of this overall system to one focusing on the use of uncertain reasoning methods in the domain of biological surveillance.

The work undertaken in this project relates to the use of methods of uncertain reasoning to classify river water quality from benthic invertebrate data, referred to in this thesis as the *direct interpretation* of the source data. While the computer software[1] developed during this project to automate this process of classifying river water quality

[1] These software systems are referred to as “automatic classifiers”.

from benthic data could be adapted to form one part of the overall BERT system, its main purpose was facilitate an experimental investigation of the use of uncertain reasoning methods in this domain. Most of the software development took place within the LEONARDO expert-system shell (Creative Logic, 1992), and could be easily integrated within the BERT decision-support system as originally conceived.

1.5 Organisation of the thesis

Biological methods of river water quality assessment and their role in river management are discussed in Chapter 2. Here the meaning of the term "water quality" as used for rivers and freshwater bodies is considered, followed by a review of the theoretical foundations of biological surveillance and its relation to chemical monitoring. The chapter includes a critical evaluation of the many schemes for reporting and summarising biological river water quality, and a discussion of the current trend for incorporating biological quality measures in British river quality surveys.

The need for representing uncertain information in human expertise is reviewed in Chapter 3. The close association between methods for managing uncertainty and those of decision making is discussed. The primary numerical unreasoning scheme, Bayesian theory, is discussed with reference to its simplified representation and to the growing interest in Bayesian networks, which represent the most powerful expression of probabilistic methods to date. Rival schemes important in expert system technology are also reviewed, with particular attention paid to the Dempster-Shafer theory of evidence, a generalisation of the Bayesian calculus, and possibility or "fuzzy-set" theory.

In chapter 4, the main knowledge elicitation work undertaken in cooperation with the domain expert is described. Benthic taxa of particular value for indicating water quality were selected, and ecological knowledge on the benthic organisms was acquired. After reviewing various techniques for eliciting measures of uncertainty, the processes used for deriving the likelihoods of occurrence of the sensors across the quality classes and for representing these as discrete probability distributions are described in detail.

The construction of benthic sample data sets from invertebrate records in described in Chapter 5. Data was classified by the domain expert in terms of the discrete biological classes, and provided the reference against which the performance of the automatic classifiers could be assessed.

Chapters 6 and 7 contain the results of the automatic classifications and analyses of the computational aspects of the use of the uncertain reasoning calculi, weighting of evidence, evidential conflict resolution and belief representation. Decision mechanisms for summarising the output from the automatic classifiers, which allow a comparison with the expert's assessment, are discussed in Chapter 6. Results from the Bayesian and the Dempster-Shafer calculus in which evidential support for the classes is represented as Bayesian belief are evaluated. The results using the Dempster-Shafer theory of evidence, in which belief is represented in two ways that are quite distinct from Bayesian belief, are analysed in Chapter 7.

In Chapter 8, the potential for uncertain reasoning methods in future work for biological surveillance is discussed. The integration of biological classification software within an overall decision support system for river pollution is considered. Requirements for the consensus of expert opinion for adopting a system of discrete biological classes, the agreement of abundance levels and sensor states, and the combination of probability distributions produced by multiple experts, are discussed. Finally, the thesis is concluded in Chapter 9.

The appendices contain material to support the main text. These include, for reference, the NWC class definitions, probability distributions (elicited and derived) of the benthic sensors, details of two Dempster-Shafer evidence combination schemes, a cross-reference table of classification experiments referenced in the text, and a comprehensive glossary and list of abbreviations.

Chapter 2

Biological Assessment of River Water Quality

2.1 Introduction

The purpose of this chapter is to consider biological methods of water quality assessment and their role in river management. The meaning of water quality is examined, and considered in its context of environmental quality, with physical, chemical and also biological dimensions. The nature and role of biological classification systems for reporting river water quality are considered in detail. Finally, the advantages of discrete biological classification are described.

2.2 River Management and Water Quality

2.2.1 Major river uses

Rivers are the most important freshwater resource, supporting a multiplicity of uses from the supply of drinking water to land irrigation and waste disposal. Rivers and the corridors of land through which they flow provide a major wildlife resource and as such are enjoyed for their aesthetic and recreational value also. Inevitably, the various uses of the river can lead to a conflict of interests. The use of river water upstream for instance, must not lead to a serious degradation of quality and quantity for those downstream. Rivers have a dual role in acting as channels of disposal for municipal and industrial wastes and supplying water: such services must be provided at economic cost, while satisfying exacting quality standards. Therefore, the management of rivers is a complex problem with social, economic and political dimensions beyond its scientific aspects (Hawkes, 1979b).

River management has been defined as "regulating water quantity or quality so as to ensure (1) the most economic use of an available resource, and (2) conservation of the natural environment" (Chandler, 1970). The second provision is seen by Chandler to be of paramount importance. Conservation of the natural environment not only ensures the continuity of flora and fauna that may otherwise be eradicated, but also allows for the provision of the various uses. Others recognise that while certain rivers should be selected for conservation, the pristine quality of many rivers must inevitably be affected by human activity; it is however in our best interests to minimise this influence (Hawkes, 1979b).

Assessments of river water quality need to embrace these dual roles of management, those of the river as a natural resource and amenity, and the several uses of the water.

2.2.2 Defining river water quality

The complexity of factors that affect and determine river water quality precludes a simple definition. Hellawell (1977) considers that water quality is "difficult to define, impossible to measure absolutely and, as an abstract concept, very subjective". Subjectivity may be avoided by relating water quality to water use, so that by adopting appropriate criteria its quality may be judged to be high or low. This however, assumes that adequate criteria for the particular use can be determined. Use-related definitions of water quality are therefore relative: water regarded to be of acceptable quality for one purpose, for instance in the cooling of power stations, may not be acceptable for another, e.g. for potable supply. Multi-purpose uses of the water further complicate definitions of water quality (Hellawell, 1978).

A river, like any freshwater body, can be characterised by its hydrological, physico-chemical and biological components, and a complete water quality assessment requires appropriate monitoring of all three. According to Meybeck *et al.* (1992):

- The *quality of the aquatic environment* can be defined as (i) a set of concentrations, specifications, and physical partitions of inorganic and organic substances, and (ii) the composition and state of aquatic biota found in a water body.

- *Pollution of the aquatic environment* means the introduction by man, directly or indirectly, of substances or energy which result in such deleterious effects as (i) harm to living resources, (ii) hazards to human health, (iii) hindrance to aquatic activities including fishing, (iv) impairment of water quality with respect to its use in agricultural, industrial and often economic activities, and (v) reduction of amenities.

Since changes to river hydrology affect plant and animal communities, such modifications could be deemed as pollution under this second definition, either as a reduction of an amenity or by harming aquatic habitats. These two definitions together therefore incorporate the three major components of freshwater quality. Descriptions of quality can include quantitative measurements (for instance, physical and chemical characteristics of the water, sediments, organic matter, and so on) and qualitative and semi-quantitative descriptions of the state of the biota (Meybeck *et al.*, 1992).

Physical, chemical and biological methods are therefore required to measure the parameters that define aquatic-environmental quality. Concern for water quality issues arises when one or more of Man's requirements of the resource are affected. Physico-chemical attributes of water (acidity, salinity, hardness, suspended solids, etc.) are of primary interest where water is abstracted for domestic or industrial supply. If the river water is used for fisheries or conservation, the main issue is the ecological health of the river. The latter is best measured by biological surveillance methods, which directly assess the state of the biota and can indirectly monitor physico-chemical characteristics of the water and the effect of physical changes to river systems.

2.3 Biological Methods of River Water Quality Assessment

2.3.1 Effects of pollutants on biological communities

According to Jeffries and Mills (1990) all pollutants are characterised by their resultant degradation of ecosystems, and a consequent reduction in the quality of the aquatic environment. These changes produce a variety of effects on aquatic organisms. Observation of the state of these organisms can provide information on a large range of water quality issues. These include the effects of discharging pollutants into rivers, assessments for particular uses (e.g. fisheries) and monitoring the effects of physical changes in the river corridor (e.g. afforestation, canalisation or impoundment).

Discharges of polluting substances in rivers have been extensively studied for their effect on biological communities. Organic enrichment is the commonest source of pollution, arising from the introduction of domestic sewage, urban runoff, industrial effluents and farm wastes into the aquatic environment (Mason, 1991). It is important for lotic waters such as rivers, owing to their use as channels of waste disposal. Pollution by organic matter is complex, involving the aggregation of various factors such as oxygen depletion, toxicity from ammonia and sulphur compounds and the presence of suspended solids (Hynes, 1960). Each of these factors can have far-reaching effects. Suspended solids, for example, often result in reduced photosynthesis due to the decreased levels of light, interference with the feeding of animals that use filtration, and degradation of the substratum, affecting all animals dependent upon this habitat (Warren and Doudoroff, 1971). Organic pollution thus has consequences for both the physico-chemical characteristics of the aquatic environment and the biota that depend upon it.

The effects of a severe organic discharge into a river have been described by Hynes (1960). The putrescible wastes are oxidised and decomposed by microorganisms that rapidly increase in number, exerting a biological oxygen demand (BOD) which results in the oxygen sag characteristic of organic enrichment. There are also marked and clearly identifiable changes in the benthic communities. Invertebrates that cannot tolerate the low oxygen concentrations are depleted and eventually eliminated, replaced by those that can benefit from the increased food and reduced competition and predation. Individuals of such species may therefore increase rapidly in numbers, while overall the numbers of species may decline, reducing community diversity. As purification proceeds, a succession of taxa dominate, in response to the nutritional and oxygen content of the water: e.g. Oligochaeta, Hirudinea and Chironomidae at different stages of self-purification after severe organic loading (Hawkes and Davies, 1971).

With toxic pollution the observed biological response is simpler: a reduction in abundance and diversity. However some species may benefit from the elimination of predators and competitors more susceptible to certain poisons. Certain taxa which are sensitive to organic pollution (for instance stoneflies) are tolerant of acid or metal pollution (Hawkes, 1964).

2.3.2 Benthic macroinvertebrates for biological assessment

2.3.2.1 Advantages of benthic macroinvertebrates

Ideally the response of the entire aquatic community to environmental stress should be studied, since all react to various kinds of pollution. (Hynes, 1960). This is unrealistic for several reasons, among them the limitation of resources and minimal knowledge of the physiological responses of most species, so that particular 'indicator organisms' - those that provide an indication of the quality of their environment - are used in practice (Hellawell, 1978). Of the major taxonomic groups used as biological indicators of pollution, including bacteria, algae, protozoa, macroinvertebrates and fish, the most popular are the benthic macroinvertebrates, those living on or in the river bed (Hawkes, 1981). Some typical invertebrates used for biological surveillance are shown in **Figure 2.1**.

The popularity of these organisms for surveillance[1] is due to several reasons. Many species have low mobility and long life spans, factors that render them valuable for finding polluting discharges and for integrating local changes in water quality over time. Qualitative

---

[1] The terms 'surveillance' and 'monitoring' are used interchangeably in this discussion, but in the literature on water quality they have precise definitions. Useful operational definitions for the assessment of the aquatic environment are: monitoring - long-term, standardised measurement, observation, evaluation and reporting of the aquatic environment in order to define status and trends; survey - a finite duration, intensive programme to measure, evaluate and report the quality of the aquatic environment for a specific purpose; surveillance - continuous, specific measurement, observation and reporting for the purpose of water quality management and operational activities (Meybeck *et al.*, 1992)

Aston University

Content has been removed for copyright reasons

Figure 2.1 Examples of benthic macroinvertebrates used for biological surveillance (Figure compiled by the author from drawings supplied by H.A. Hawkes) sampling is easy due to their ubiquity and abundance, while the availability of good taxonomic keys for the macroinvertebrates makes identification less onerous than that for other groups. Specific ecological requirements are generally well known, as are their responses to different types of pollutant (Metcalfe, 1989; Hawkes, 1980; Hellawell, 1978).

2.3.2.2 Effect of local conditions

Against these advantages are the difficulties involved in quantitative sampling of the benthos, the seasonal occurrences of certain major phyla (e.g. the insecta), and the influence of physical factors such as the velocity of the stream current and the condition of the river substratum, both of which strongly influence the nature of the biocenoses. These physical factors are normally regarded as "non water-quality criteria" with respect to the various uses, and illustrate the importance of accounting for local conditions when using the biota for quality assessment (Hawkes, 1981). Although different river habitats, distinguished largely by these physical site parameters will support dissimilar biocenoses, the water from these different habitats may have identical physico-chemical characteristics, and may therefore be suitable for the same range of uses after abstraction.

This qualification is clearly important in the biological surveillance of pollution when interpreting the state of the benthic communities, which will differ according to river biotope. Sensitivity to local conditions may be viewed as a limitation to the method: in practical terms it means that due recognition must be paid to their influence (Hawkes, 1978; Hynes, 1960; Friedrich *et al.*, 1992). Riffles, generally fast-flowing waters with stony, eroding substrata, are the preferred locations for benthic surveillance of river water quality. Riffle communities are more responsive to changes in water quality and are more readily sampled.

By recognising the differences in communities arising from the natural characteristics of the biotopes, the quality of the aquatic environment as indicated by the biota may be used as a pointer to the quality of the water itself, and to the range of uses for which it is suitable.

2.3.3 Biological surveillance and chemical analysis

Reference has been made to a perceived limitation of biological surveillance, namely the sensitivity to local conditions. A more genuine limitation of biological surveillance using macroinvertebrates lies in the inability of the method to detect harmful pathogens or trace chemicals, which require specific bacteriological or chemical tests (Hawkes, 1974). Thus, water from a river that is of good environmental quality may be suitable for most purposes, but not necessarily suitable for abstraction for potable supply. Chemical tests are also required to identify polluting substances exactly, although biological methods can distinguish between organic and toxic pollution. Effluent quality too requires information on allowed concentrations, which cannot be directly measured by biological methods. Chemical procedures are therefore necessary for several areas of pollution and water-quality assessment. Their historical importance in the field of sewage treatment, dating from the work of the Royal Commission on Sewage Disposal in the early part of the century, probably accounts for their predominance in the practice of water-quality assessment (Hynes, 1960).

Physico-chemical tests are routinely used in pollution control surveys, involving an appreciable range of criteria (Hawkes, 1974). Of the many analyses employed, certain tests relate particularly to sewage treatment, notably biological oxygen demand, dissolved oxygen content of the water, and measurements of phosphates and nitrogen in various forms. These criteria still form the basis of river water quality classification in England and Wales (National Rivers Authority, 1991a). Traditional chemical monitoring programmes are likely to continue to play a major role in the management of water quality (Martin, 1993).

It is generally recognised that biological and chemical methods are complementary in this field (Hawkes, 1979b). Despite this however, and any intrinsic limitations of biological surveillance, there remain several aspects of water-quality assessment in which biological methods excel, and in which chemical tests are deficient. The latter have been criticised in that they yield only indirect information on the state of the biota after stress from pollution, an essentially biological phenomenon. In assessing the damage to aquatic life by toxic pollution for instance, measurements of toxicity levels commonly use standard organisms (such as *Gammarus pulex*), kept in carefully controlled laboratory conditions and subject to controlled doses. Problems arise from applying the results of such tests to biotic communities in the environment (Jeffries and Mills, 1990).

Biological surveillance directly examines the effects of water quality changes on aquatic life. The biota act as continuous sensors, reacting to past and present fluctuations in quality parameters and integrating changes over time. Because of this, the frequency of sampling for pollution surveys can be reduced. For chemical tests, sampling frequency must be much higher, since they measure conditions at the time of testing and can miss extreme conditions such as discharges of pollutants, if these happen to occur between samples. Effective chemical testing, with the associated statistical processing of data, involves a considerable workload for water authorities (Hawkes, 1979b). Biological methods may detect unknown or unusual pollutants that would otherwise be overlooked by chemical tests, which, on economic grounds, can only deal with a small number of determinants.

For these reasons some authors have suggested that biological methods should assume more prominence in water quality assessment, or even form the backbone of surveillance programmes (Hynes, 1960; Hawkes, 1979b). In fact, these methods are now an integral part of water quality assessment (De Pauw and Hawkes, 1993).

2.4 Classification of River Water Quality

2.4.1 Purpose of data reduction

The methods of biological surveillance can be used in several applications, from general environmental surveys to the monitoring of the effects of specific discharges or changes to river hydrology, for example. The purposes of the particular survey largely determine the sampling effort and therefore the amount of resulting data, which then require processing and interpretation. Biological data from surveillance programmes is reduced or condensed to summary form suitable for decision-makers (usually non-biologists) who have responsibility for water-quality management at regional or national levels (Mason, 1991; Hellawell, 1978).

Certain authors have doubts concerning the efficacy of biotic indices for indicating biological water quality (for instance Mason (1991) and Hynes (1970)). Nevertheless, the existence of numerous schemes devised for summarising biological water quality suggests that they have a useful role in water quality assessment and management.

2.4.2 Evolution and design of biological assessment methods

The various systems have been extensively reviewed by Hellawell (1977, 1978, 1986), Hawkes (1977, 1979a), Washington (1984) and Persoone and de Pauw (1979). Recently, Metcalfe has presented both the historical development and the current state of European bio-assessment systems that use the benthic macro-invertebrates, identifying three main categories from the huge number of indices now in existence: the saprobic, diversity and biotic approaches (Metcalfe, 1989). This categorisation forms the basis of a recent comparison of biological assessment methods (de Pauw and Hawkes, 1993), and is adopted here for the following review, along with a further category of statistical methods. Given the enormous quantity of schemes in existence, the review is restricted to a small subset of the most important schemes.

2.4.2.1 Saprobic approach

Different groups or species show different responses to pollution, with each species preferring a definite range of environmental conditions: a phenomenon that facilitates the biological classification of rivers (Chandler, 1970). For some species this range can be very narrow. One of the earliest biological classification schemes for surface waters was the saprobic system introduced at the turn of the century by Kolkwitz and Marsson (1908, 1909). This was based on the idea of zones of organic enrichment of a river, each of which had an associated aquatic community indicative of varying degrees of water quality. The system suggests that the zones correspond to different levels of the self-purification or saprobity of the water: the polysaprobic zone represents the severest level of pollution immediately downstream of the organic load, with recovery occurring in the α-mesosaprobic followed by the ß-mesosaprobic stage. The oligosaprobic zone represents full recovery to the "appropriate, natural community for the river channel as it stands downstream" (Jeffries and Mills, 1990).

Zones are characterised by certain indicator species, chemical conditions and the nature of the water. Since the indicator species are associated with particular zones, comparison of the species list from a sample with the indicator species occurring in the four zones allows one to classify the river water into quality categories (Friedrich *et al.*, 1992). The Saprobic Index based on this classification allows the species lists to be condensed to a single number showing the saprobic zone at the sampling site.

Variations of the index exist due to several workers (Pantle and Buck (1955), Zelinka and Marvan (1961) and Sládeček (1973)). Of particular significance is the concept of "saprobic valency" introduced by Zelinka and Marvan to represent the likelihood of species occurrence across five saprobic zones (the additional xenosaprobic zone representing a higher quality than oligosaprobic). An expert (or a group of experts) allocates the valencies for particular species - a subjective exercise relying on experience of the ecological ranges in which they occur.[2]

Table 2.1 Examples of saprobic valencies of benthic invertebrates

Aston University

Content has been removed for copyright reasons

Source: Sládeček (1973)

Table 2.1 shows examples of valencies for the five saprobic zones of selected benthic invertebrates. Note that certain taxa occur across three or more zones, while others are "concentrated" in two adjacent zones. The latter are considered of high "indicator value". Sládeček (1973) incorporated the concept of valency into "s values" representing the preferred saprobic zone of a species, and utilised abundance-levels and indicator value for the Extended Saprobic Index:

$$S = \frac{\sum_{i=1}^{n} s_i \cdot h_i \cdot g_i}{\sum_{i=1}^{n} h_i \cdot g_i} \qquad \textbf{(2.1)}$$

Here, S is the overall saprobic index, n is the number of species in the list, h_i is the abundance-level of species i, s_i is a value corresponding to the preferred saprobic zone of the species, and g_i is a weighting factor representing the value of the species as an indicator.

The saprobic system is demanding in terms of required taxonomic information. Organisms must be identified to species level, and their occurrence in each of the river classification zones for a particular region must be known so that they can be assigned to a preferred saprobic zone for purposes of calculation. The system is clearly designed for detection of organic pollution using indicator species: community responses such as loss of diversity, which are characteristic of toxic pollution, are not directly accounted for.

[2] See Sládeček (1973) for an objective method for assigning saprobic valencies.

However, in spite of severe criticisms of the saprobic system (e.g. Hynes (1960)), it has found favour in Continental Europe, forming the basis of several indices. Recently, it has been extensively revised to incorporate practical experience and the results of physico-chemical analyses. It now forms part of a standardised and integrated system of water quality classification in Germany (Friedrich, 1990).

2.4.2.2 Biotic approach [3]

Trent Biotic Index

The Trent Biotic Index (TBI) is based on the number of defined groups of taxa present in relation to six "key" organisms: Plecoptera, Ephemeroptera, Trichoptera, *Gammarus*, *Asellus*, tubificid worms and/or red chironomid larvae. As with the saprobic system, this behaviour corresponds to the effects of increasing organic enrichment (Woodiwiss, 1964). The TBI incorporates a measure of diversity (via the number of "groups" present, and by the numbers of species within the identified "key" organisms). However, it does not account for abundance-levels either within the groups or within the key organism species. Thus, the unexpected presence of a single individual of a sensitive species can affect the index disproportionately (Chandler, 1970; Mason, 1991).

Chandler Score

In addressing this limitation Chandler introduced a score system in which five abundance-levels were identified: these were used to weight the score of indicator species-groups, with pollution-sensitive species scoring higher than those more tolerant to organic enrichment, and with increasing abundance within groups receiving greater or lesser weight depending on tolerance to organic pollution. Measures of absolute abundance were rejected due to associated sampling errors; subjective assessments of abundance however were deemed to be of sufficient accuracy when several samples are taken from the same station during the year. With the incorporation of abundance-levels, the Chandler score indicates the biological condition of the river via the diversity and abundance of the invertebrate fauna. Unlike the TBI however, the score is continuous: it is claimed that these features clearly illustrate differences of diversity and abundance between stations, and eliminate "borderline" class assessments.

---

[3] Strictly, the saprobic system is a biotic approach. The categorisation adopted here is that of Metcalfe (1989) and de Pauw and Hawkes (1993).

BMWP Score

The TBI and Chandler score have been largely superseded in Britain by the BMWP (Biological Monitoring Working Party) score. This was designed to be simple to use, with a requirement of taxonomic identification only to family level. It was to be a "surveillance tool" allowing the monitoring of temporal trends in biological water quality at a site, as opposed to regional spatial trends (Chesters, 1980). However, its use in the national river survey of 1980 for England and Wales inevitably invited comparisons between rivers achieving different scores, in spite of zonal or geographical dissimilarities (Hawkes, 1978).

Groups of invertebrate families are assigned points out of 10 depending on their intolerance to organic pollution. The score is calculated by identifying animals to family level and accumulating the equivalent points for each 'scoring' taxon in the sample. If a BMWP family is already represented in the sample score the further occurrence of taxa belonging to that family does not add anything to it. Furthermore the most tolerant taxon within the scoring group is selected for awarding points as a means of militating against over-optimistic assessments of quality.

The BMWP score is dependent on sample size and therefore sampling effort. By dividing the score by the number of scoring taxa, the ASPT (Average Score per Taxon) is derived which is independent of sample size and less influenced by seasonal factors which can influence the BMWP score. Metcalfe (1989) reports a study by Murphy (1978) in which BMWP score and TBI values were reduced at headwater sites, which, as will be clear from the discussion on diversity, may more accurately reflect the physical conditions present rather than water quality. By converting to ASPT, the values obtained were then commensurate with the high quality water of such streams.

The ASPT is less sensitive to toxic pollution. However, the use of both scores in tandem may have some merit, by indicating the type of pollution. Low ASPT and BMWP scores occurring together point to organic enrichment, whereas low BMWP coupled with a high ASPT suggests physical or toxic environmental stress (de Pauw and Hawkes, 1993).

The use of the BMWP score as a spatial indicator of quality trends is problematic where markedly different faunal communities occur due to particular geographical or physical conditions. The benthic communities of upland rivers for instance, are inherently different from lowland rivers, even though the water may be of comparable chemical quality. A low-scoring lowland river might be thought of as being "biologically inferior" to a high-scoring upland river, even though the communities for both types may be "appropriate and natural". According to de Pauw and Hawkes (1993) "river benthic invertebrates are only of value as indicators of river water quality when considered in the context of the biotope in which they were found."

Lincoln Quality Index

The specific problem of geographical influences on the BMWP score was addressed by Extence *et al.* (1987) who wished to use biological methods directly for water quality management. They devised the Lincoln Quality Index (LQI), based on the BMWP and ASPT scores, with the intention of producing a simple index which provided freshwater quality data to operations managers in the Anglian Water Authority. The impetus to derive a new index came from such considerations and the desire to combine information provided by both the BMWP and ASPT scores. Essentially, this information is represented in a way that is readily understandable, and which avoids the need for qualifying explanation.

The LQI requires that sample sites are pre-classified as either habitat-rich riffles, or habitat-poor riffles/pools, the latter receiving enhanced ratings to 'compensate' for their inherently sparser communities. Standard sampling techniques are used, from which the BMWP and ASPT scores are calculated. Ratings for bands of BMWP and ASPT scores are assigned according to the site type: the rating values and the score-bandings were derived from analysis of the data in the Anglian region. **Table 2.2** shows the standard ratings X and Y (parameters for the calculation of the LQI) for habitat-rich riffles; enhanced ratings are

Table 2.2 Lincoln Quality Index ratings for BMWP and ASPT scores

Quoted ratings X and Yare for habitat-rich riffles. See text for explanation. Source: Extence *et al.* (1987)

used for habitat-poor riffles and pools.

The overall quality rating is the average of X and Y, which is then used to derive the LQI itself. An overall rating of 6 or better yields an LQI of A (excellent quality); ratings of 1½ or lower are LQI H or I: very poor quality. The X and Y ratings chosen have the effect of limiting high ASPT and BMWP scores. If low BMWP and high ASPT (or vice versa) occur together, the higher rating X or Y promotes a correspondingly higher LQI.

This approach is novel in that it attempts to directly account for faunal differences in habitat type. A similar idea was considered for the BMWP system, but not actually incorporated (Hawkes, 1979a). According to Mason (1991), the LQI system "provides a real attempt to make good management use of routine biological surveillance data".

2.4.2.3 Diversity approach

In general, good quality waters will support diverse communities, including a proportion of species sensitive to pollution. Grossly-polluted communities tend to have low diversity, with a high proportion of pollution-tolerant taxa. Diversity indices attempt to quantify this phenomenon mathematically by combining data on species abundance into a single number.

There are several different formulations of diversity indices, extensively reviewed by Washington (1984) and Hellawell (1977). Early formulations were found to be unsatisfactory due to their dependence on sample size and their failure to express the relative importance of different species within the community, unlike those derived from information theory. The use of information theory indices in the analysis of biological water quality was advocated by Wilhm and Dorris (1968) who considered them to be superior in expressing community diversity. Wilhm and Dorris derive a diversity index D:

$$D = -\sum_i \frac{N_i}{N} \log_2 \frac{N_i}{N} \qquad \textbf{(2.2)}$$

where N_i = number of individuals in the ith. species, N = total number of individuals in all s species.[4] Other current diversity indices are reviewed by Washington (1984).

Communities may be considered to be unstressed if they exhibit high diversity, in which there are a large number of species each represented by few individuals (Hellawell, 1986). A change in the benthic community's diversity index will in theory signal a change

[4] This formula quoted by Wilhm and Dorris is in fact an approximation of Brillouin's diversity index, but is usually referred to in the literature as the Shannon-Wiener (or Shannon-Weaver) index.

in water quality, since the structure (as summarised by a diversity index derived from information theory) will be altered by environmental stress due to pollution. However, the practical application of diversity indices for water quality assessment requires caution. While a high diversity index generally indicates good quality water, the converse may not be true: a low diversity index may be due to other factors apart from poor water quality. Adverse physical conditions in fast-flowing streams will reduce diversity, often to very low levels, even though chemical water quality may be good.

Diversity indices have been described as insensitive to changes in water quality, and exhibiting poor site discrimination (Metcalfe, 1989). Green (1979) has more fundamental objections. He argues that there is not a meaningful relationship between diversity indices and community structure, and questions whether "a diversity index is the most efficient and biologically interpretable way to summarize biological data". Simple indices such as number of species are considered more biologically meaningful measures than the complex information-theory indices (Green, 1979).

The idea of environmental stress for communities affected by organic enrichment is problematic, since certain species will benefit from the increased nutrients and in these conditions cannot be considered under stress. However it is considered that diversity indices may be used to measure physical and toxic pollution because of the general stress imposed on the entire community (Hawkes, 1977).

2.4.2.4 Statistical methods

Classification and ordination methods derived from multivariate statistics have been applied to benthic community data for the purposes of environmental surveillance. A common technique is that of cluster analysis performed on Jaccard similarity coefficients to identify 'affinities' between invertebrate groups (Jaccard, 1912; Hellawell, 1977). By associating these groups with site physical and chemical data, causal factors that determine the distribution of the invertebrates may themselves be recognised. Brooker (1984) considers that these methods offer a powerful means of identifying environmental change. One of the more important developments in river water quality assessment in Britain in recent years has been the use of such techniques to predict the composition of benthic communities from the physical and chemical characteristics of freshwater sites, and to classify the sites from their fauna. Because of its significance, it is considered here in some detail.

The work has been described in several papers, including Armitage *et al.* (1983), Wright *et al.* (1988), Armitage (1989) and Wright *et al.* (1989), while a good overview is given by Metcalfe (1989). It relates to an extensive research programme which was carried out over several years by the FBA, and subsequently the IFE, in which a large number of unpolluted sites covering the whole of Britain were surveyed for macroinvertebrate fauna. The result of this work was a comprehensive database of taxa lists and environmental and physical parameters for the sites.

The sites were clustered into site groups from their constituent fauna, using a multivariate technique known as TWINSPAN (Hill, 1979). Once the groups were defined in this way, discriminant analysis was used to differentiate between the groups of the basis of their environmental characteristics, which were chosen as the discriminating variables. Statistical tests showed that there was a good fit between the environmental features and the site groupings (Wright *et al.*, 1989).

The analysis will classify new sample sites into one of the groups. The frequency of occurrence of each species within each group is determined by examining the taxonomic database to discover the proportion of sites at which a given species or family occurs. This, combined with the probability of the new site belonging to a particular group, yields the probability of capture of species at the site, under the standardised sampling conditions. Thus, the method involves the classification of new sites and the prediction of faunal composition.

A software package called RIVPACS (River Invertebrate Prediction and Classification System) has been written to automate this analysis (Wright *et al.*, 1989). It accepts physical and chemical variables for a freshwater site and outputs the probability of the site belonging to one or more of the groups, and the predicted fauna in order of decreasing probability. Initially, 28 environmental parameters were used for prediction, but these were reduced, by stepwise multiple discriminant analysis, to subsets of the original. Some of these reduced sets consisted of combinations of physical and chemical variables; others used solely physical parameters in an attempt to predict benthic communities at sites possibly subject to chemical pollution.

The programme has been extended to produce benthic community predictions at family-level, and to predict BMWP score, number of scoring taxa and ASPT using multiple regression equations (Wright *et al.*, 1989). With this predictive ability, RIVPACS highlights deviations from reference conditions as indicative of possible environmental stress. This property has been used in a practical application of the biological surveillance of river water quality. For instance, in Wright *et al.* (1988), the authors used the system to detect organic pollution in a river by monitoring three sites, one upstream and two downstream of a sewage treatment plant. The environmental parameters for the sites generated the expected BMWP families, so that the ratio of observed: expected number of families were derived for each station. The exercise was repeated for species prediction and ASPT. It was found that the ratios of the derived quantities decreased from the upstream control site to the site downstream of the organic discharge, recovering at the site furthest downstream. Comparison of predicted and actual species lists was also of particular value in emphasising the loss of taxon richness.

The introduction of RIVPACS has had an important bearing on river quality classification in Britain, as will be discussed in the next section.

2.4.3 Discrete biological classification schemes for river water quality

2.4.3.1 Discrete and continuous schemes

By definition "to classify" means to "arrange in classes; assign to a class" (Allen, 1990). In his article on river zonation Hawkes considers that classification "enables items to be placed into classes the members of which have definable characteristics in common" (Hawkes, 1975). Ideally, the classification system will encompass all the likely conditions to be encountered in its domain, for example, the range of water qualities in Britain. The number of classes in the scheme must be chosen with some care. Too few classes results in a coarse system that will be insensitive to important changes in quality and therefore of limited practical use. With a large number of classes (and therefore more class boundaries), the chances of misclassification error increase (due to sampling error, for instance).

In contrast to the discrete classification approach, water quality may be measured as a continuous rather than a discrete variable. Both variants have been used in the quality classification of rivers, with some systems incorporating aspects of both. For instance, the saprobic systems are based on a notion of "zones of organic enrichment", with the presence or absence of indicator organisms coupled with chemical characteristics of the site determining the zone. Additionally however a single figure, the saprobic index, may be derived. One assumes that both the discrete zone and the numerical value were useful for summarising river water quality using the saprobic approach.

Other schemes adhere to one classification mode. The Trent Biotic Index (Woodiwiss, 1964) was designed partly to eliminate unbounded estimations of quality at the top end.[5] Woodiwiss considered this to detract from the ability of the Index to communicate information on river quality to managers or non-biologists effectively. However the Index, which originally consisted of 10 classes was extended later to 15 to increase the sensitivity of the scheme to the good-quality range of waters (Hawkes, 1979a). European derivatives of the TBI are also discrete classification-based schemes, the Belgian Biotic Index (BBI) being related to quality classes within very narrow bands (de Pauw and Hawkes, 1993). Similarly the Lincoln Quality Index (Extence *et al.*, 1987) is expressed as a class letter (A to I).

Continuous quantification schemes such as the score systems of Chandler and the BMWP both incorporate indirect expressions of diversity, in that theoretically unstressed communities will consist of many species, leading to potentially very high scores. Diversity indices, specifically designed to express these attributes of biological community structure, are also continuous measures.

It may be significant that bio-assessment schemes that have been most criticised in the field, apart from any theoretical reservations, are those that are not class-based, such as the BMWP score and diversity indices. It would appear that discrete-classification schemes are more easily understood and allow better communication of biological information for management purposes. In this context the discrete classification of river quality is perceived as a valuable exercise.

2.4.3.2 History of discrete classification schemes in Britain

Biological classification was used as part of the first official national river survey of England and Wales in 1970, to supplement the established chemical classification. For the biological assessment, waters of four distinct quality classes (A to D) were defined in terms of the fauna they were likely to support, with class A being of the highest quality. A comparison of the biological and chemical classifications "failed to show the agreement ...which was considered necessary before biological data could be used to report on river water quality on a national scale" (National Rivers Authority, 1991b).

However, the survey illustrated an important aspect of biological surveillance that needs to be accounted for in any system employed on a wide scale. The biological survey was carried out for rivers that have markedly different physical characteristics, particularly

---

[5] The author interprets this as a reference to continuous schemes such as the Chandler or BMWP scores. In practise however, these scores have do have fixed upper limits.

current velocities. Thus the slow-flowing rivers of East Anglia were predominantly graded to be of lower biological quality than their corresponding chemical class, by virtue of the fauna chosen to define the biological classes (Department of the Environment, 1972). Invertebrates characteristic of the rhithron zone, and associated with high quality waters in the adopted classification, are less unlikely to be found in the potamon waters of the low-lying East Anglian rivers (Hawkes, 1975).

A discrete biological-classification scheme was used by Severn-Trent Water for their 1987/88 report on the biological quality of rivers in their region. As with the 1970 national survey, the river classes were defined in terms of the biota. For example, class 1A (biological) is defined:

> Good water quality with high dissolved oxygen and water velocity giving rise to habitats suitable for a diverse fauna with stoneflies, mayflies and caddis flies in high numbers.

The reported biological classes are then used to detect deviations from the chemical classes. However, biological sampling data is summarised using the BMWP and ASPT scores, which are then categorised into six grades labelled Unsatisfactory to Excellent (Severn-Trent Water, 1988). A similar biological class system was employed by Yorkshire Water Authority (YWA)[6] and its successors, referring to classes B1a to B4, parallelling the National Water Council (NWC) classes.

2.4.3.3 Introduction of NWC classes

The NWC classes introduced in the late 1970s were to provide both an absolute measure of river quality and a means of reporting trends. The classes were defined in terms of the levels of determinands needed to protect the more important uses of the river, such as fisheries and abstraction for potable water supply (see Appendix A1). As more uses were introduced, each accompanied by standards and EC-directives, there was growing concern that the ability to measure absolute quality would disappear, since not all uses apply to all rivers. In response to this concern a working group within the water industry recommended that a classification scheme be specifically defined to measure absolute quality: this has been accepted by the NRA (National Rivers Authority) who have proposed a new general

---

[6] Reference will be made to this body throughout this thesis, although it no longer exists, and also to the National Rivers Authority (Northumbria and Yorkshire Region), the successor of the Yorkshire Water Authority. The NRA is itself now part of the Environment Agency. Similarly, the National Water Council no longer exists.

classification system for rivers, besides new use-related classes that will replace the old NWC system. These schemes follow from the imposition of Statutory Water Quality Objectives under the 1989 Water Act for specified bodies of water in England and Wales (National Rivers Authority, 1991a). Clearly some merit is seen in reporting river quality from two distinct but complementary perspectives: that of the "health" of the river, in terms of a general class, and that of the uses for which the river can be put.

2.4.3.4 Recent developments in River Water Quality classification

The NRA's proposed use-related classes are defined in relation to recognised uses of the water, such as abstraction for potable water supply, water contact activity, fisheries, harvesting of fish, abstraction for industrial and agricultural use. Of particular interest is the introduction of two new uses that relate to the integrity of aquatic ecosystems: a "general ecosystem" for all controlled waters and a "special ecosystem", for designated sites which "have special requirements for nature conservation" (SSSIs, for example). Here the conservation of the water for aquatic life becomes an amenity, for which quality standards may be employed to judge whether a stretch of water is suitable for the stated use. It is noteworthy also that, except for the two ecosystem uses, all other uses are associated with classes defined entirely by physico-chemical determinants (National Rivers Authority, 1991b).

The general ecosystem classification evolved in part from the extensive work carried out by the FBA (and subsequently the IFE) in the development of RIVPACS. The results of this work were used by the NRA to make a biological assessment of Britain's rivers in the 1990 national survey, during which physico-chemical data and biological samples were collected for nearly 9000 sites. Twelve environmental variables (which previous testing had shown to yield good predictive ability) were used to predict BMWP, ASPT and total numbers of taxa: these could then be compared with the actual values obtained from sampling. For each of the three parameters (BMWP score, taxa number and ASPT) a ratio of observed to expected results was calculated to produce three Ecological Quality Indices: EQI (BMWP-score), EQI (number of taxa) and EQI (ASPT). Ranges of the three EQIs were then defined corresponding to biological classes A to D (Sweeting *et al.*, 1992). Using this system, a biological classification of Britain's rivers was conducted.

It is recognised however that biological classification and chemical classification produce two quite different measures of water quality. As a means of combining the two into a single system, it has been suggested that the EQI (ASPT) could act as a "biological override" to the chemical classification. This would allow a chemically-classed river to be upgraded or downgraded depending on the biological quality suggested by the EQI (ASPT) (National Rivers Authority, 1991b).

Since then, the British Government has issued its own proposals building on those outlined by the NRA. Draft Regulations have been produced to set up a new system of classifying river quality in England and Wales (Department of the Environment, 1992, 1993). The regulations deem that the NRA's proposed "general ecosystem" Use class does not meet certain criteria of simplicity, efficiency and enforceability. Instead, the cornerstone of the new Use Classes system is the introduction of the "Fisheries Ecosystem" classification for general water quality, allowing the setting of targets for river quality objectives for fish and aquatic ecosystems. Many determinants used to define the NWC classes are employed, along with additional parameters such as dissolved copper and zinc - six classes FE1 to FE6 are thus defined.

A separate General Quality Assessment (GQA) for the five-yearly national surveys is proposed to give "an objective measure of how river quality is changing over time". The Government considers that while biological monitoring techniques are acknowledged to be an important part of the GQA, they are not sufficiently developed to allow inclusion in the statutory scheme. The GQA would include four types of assessment: a chemical classification A to F (directly comparable to the Fisheries Ecosystem Use Classes), a biological classification (A to E) based on the EQI, and similar classes for aesthetic and nutrient status. Therefore the current position is to view biological assessment as an important tool for water quality monitoring, but given the fact that there is not a clear relationship between chemical and biological parameters, the chemical and biological assessment of river quality should be carried out separately and in parallel, rather than to use biological overrides in chemical classification (Department of the Environment, 1992).

2.4.4 Critique of Biological Classification Schemes

Both biological and chemical analyses are required to monitor the quality of the aquatic environment, and are therefore complementary techniques. Surveillance of the biota reveals direct information on the biological quality of that environment, and indirect information on the physico-chemical characteristics of a major component: water itself. Chemical and biological sampling and the subsequent processing of the data commits an appreciable amount of water authorities' time and resources. Therefore examining current classification methods is reasonable, and to ask if they use the available data to maximum effect. This is particularly true for biological methods, since some biologists still view with suspicion the necessary data reduction that occurs in producing summary reports despite arguments for biological input to decision-making. As Extence *et al.* (1987) have argued, a quality index should "utilise as much information as possible from a sample".

The bio-assessment methods considered can be criticised on (i) their underlying philosophy, (ii) the physical design of the classification scheme and (iii) their use of the sample data. From within the first viewpoint the schemes could be divided into two broad categories: those based on the saprobic/biotic approach, in which expert knowledge on key benthic organisms have a central role in their formulation, and the statistical/community-structure approach that effectively seeks to eliminate subjectivity as far as possible. These attributes of subjectivity or objectivity (or the lack or them) associated with these schemes have each been used to criticise those in opposing categories. For instance, diversity indices devised as quantitative measures of community structure (developed in response to perceived limitations in subjective interpretation) have been faulted in respect of the fact that they do not use autecological knowledge (Hawkes, 1979a).

Biotic systems are useful as indices of organic pollution, provided due account is paid to local conditions at the sample site. Their success is perhaps surprising given the simplicity of their formulation, a consequence of the requirement that these systems should also be easy to use. An overview of important limitations of biotic indices has been given by Walley *et al.* (1992a). In this section those limitations relating to the use of data are considered in some more detail here.

The significance of varying levels of abundance of individual species or taxa is not acknowledged by some schemes. The popular BMWP score for instance takes no account of taxa abundance (unlike the earlier Chandler score) and is affected by sample size and seasonal factors. Abundance with each group is not accounted for by the Trent Biotic Index, although it does refer to absent species.

The Lincoln Quality Index (Extence *et al.*, 1987) is essentially a reworking of the BMWP score and ASPT, although one that incorporates the effects of habitat differences on taxa richness. Extensions to the saprobic system due to Sládeček (1973) and Zelinka and Marvan (1961) are significant in their acknowledgement that biota can exist across a range of water qualities (the saprobic zones), with zonal preferences represented by saprobic valency. In this respect the extended saprobic system stands apart from other bio-assessment methods.

Diversity indices based on information theory are mathematically sound but their acceptance as measures of environmental stress under organic enrichment is not universal, although they appear to have validity for conditions of toxic or physical pollution. Perhaps the most promising of the quantitative approaches is the advanced statistical model of RIVPACS, which has considerable potential for the biological surveillance of river water quality (Wright *et al.*, 1988). A 'black-box' model such as RIVPACS cannot be interrogated for cause and effect, although many associations between the variables can be explained (Armitage, 1989). Its role as a decision-support tool for river water quality has yet to be fully exploited, although undoubtedly it has promoted the use of biological methods in this area.

The design of classification systems was also examined noting that, within Britain and Europe, discrete-classification schemes have certain advantages as descriptors of chemical and biological dimensions of river quality when compared with continuous score systems.

2.5 Summary

This chapter has examined in detail the definitions of river water quality and pollution, and reviewed the theoretical foundations of biological surveillance and its relation to chemical monitoring. The many schemes for reporting and summarising the biological dimensions of river water quality have been critically evaluated, and current developments for incorporating biological quality measures in British river quality surveys have been described.

Classification schemes are useful in reporting regional or national trends in river quality, and as such have a role to play in the management of water resources. As quality control standards extend beyond regional and national boundaries to pan-national levels (e.g. European Union directives on surface water and ecological quality) the means by which water quality is measured and reported will assume greater importance. Without this concise and summarised data the detection of temporal and spatial trends in surface water quality would be lost in the mass of chemical and biological data collected annually from thousands of samples.

Rivers, our most important freshwater resource, provide for a range of needs, in which abstraction for supply and waste disposal sit alongside recreation, amenity, fisheries and conservation. Currently, the ecological quality of surface water is being promoted as a proposed directive from the European Union, one that takes into consideration the "quality of the shore, the banks, the sediment and the environment around rivers" and the water itself (Collins, 1993). Increasingly, water quality issues are viewed within the context of the overall quality of the aquatic environment. This can be monitored via three principal media: the water itself, particulate matter (organic and inorganic) and living organisms. Of these, benthic invertebrate communities are the most widely-used for assessing biological quality.

In England and Wales discrete classification systems have been used for reporting chemical water quality and will be employed for proposed use-related classes. Although discrete biological classification has, and continues to be used by certain water authorities such schemes have not formed part of national river surveys since 1970. The Biological Monitoring Working Party in developing its score system could not at the time recommend the adoption of a biological classification scheme (Department of the Environment, 1980). Currently therefore, there is not a nationally-recognised biological classification system in Britain.

Nonetheless it is argued here that given the benefits of a discrete class-based system for reporting river water quality, a discrete biological classification scheme could provide an effective means of reporting biological river water quality. Since chemical and biological methods are accepted as complementary in the field of river water quality control, it seems appropriate that the biological classes mirror the current NRA classes for river water quality, which use chemical determinants.

Chapter 3

Management of Uncertain Reasoning

3.1 Introduction

The purpose of this chapter is to review various procedures for handling uncertainty, and emphasise the close association of uncertain reasoning with decision theory. After an overview of the various methods, in which it is argued that numerical schemes are required to deal with uncertainty adequately, the Bayesian and Dempster-Shafer calculi are described in some detail. A rationale for focusing on these two procedures is presented. An important alternative theory to rival to the Bayesian and Dempster-Shafer calculi is presented: approximate or "fuzzy" reasoning. The chapter concludes by comparing the knowledge-based methods of uncertain reasoning with the artificial neural network paradigm.

3.2 Uncertain Reasoning and Decision Theory

Reasoning is closely allied to the process of decision making. If our knowledge about some problem is incomplete, deciding some course of action may be difficult or impossible, or we may make bad decisions. In practice, human decision makers constantly deal with inadequate information when making decisions: a condition that we call "uncertainty". To manage this problem, human beings or automated reasoning systems must therefore reason under uncertainty (Giarratano and Riley, 1989).

Uncertain reasoning has attracted much attention recently regarding its use in artificial intelligence, a field of study concerned with performing computational tasks that apparently require human intelligence (Tanimoto, 1990). Artificial intelligence encompasses many areas of study, from robotics, artificial neural systems, natural language processing to expert systems, which in the mid-1980's were still described as its "most visible and fastest growing branch" (Bonissone and Tong, 1985). Since expert systems are computer programs that emulate the human expert's ability to decide (Giarratano and Riley, 1989), much of the following literature review on uncertainty and uncertain reasoning deals with its management within expert systems.

Definitions of an expert system include "... a computer system that is designed to help people with tasks involving uncertainty and imprecision, and which require judgement and knowledge" (Hart, 1989) or simply "machines which reduce uncertainty" (Graham and

Jones, 1988). Henrion *et al.* (1991) in a review of decision analysis and expert systems, have argued that both disciplines share common goals. Both seek to improve human decision making by formalising knowledge so that it is open to automated reasoning. It is this aspect of uncertainty management rather than the structure and operation of expert systems that primarily concerns us here.

Uncertainty in reasoning systems arises from four main sources: unreliable information, imprecise descriptive languages, inference with incomplete information, and the aggregation of information from multiple sources (Bonissone, 1987). For the first source, uncertainty may be present in the data due to human error or errors of measurement, or may occur due to weak implications between a rule's premise and its conclusion. In this case, the degree of correlation is normally quantified by a numerical factor, as will be seen later. The second source arises from the ambiguity of natural language and the difficulty of formalising rules from it. Approximate reasoning derived from fuzzy logic offers a partial solution to this problem. The fourth type occurs when knowledge from multiple experts or sources is combined. Such knowledge may be contradictory, so that the system requires some mechanism for resolving conflict.

3.3 An Overview of Uncertainty Management Schemes

Schemes for dealing with uncertainty in reasoning systems may be divided into two broad categories: numeric and non-numeric. Numerical approaches include Bayes' theorem and its modified odds-likelihood formulation, confirmation theory as used in MYCIN[7] (Shortliffe and Buchanan, 1985), Dempster-Shafer theory (Shafer, 1976), and fuzzy reasoning (Zadeh, 1988). These methods allow the use of a calculus to combine and propagate the evidence through the reasoning process, and provide a mechanism for decision making.

Non-numeric methods such as Cohen's theory of endorsements (Cohen, 1985) are employed when uncertainty arises from incomplete information. They are considered inadequate to handle imprecise information because they lack measures by which confidence levels can be assigned (Bonissone, 1987). Shafer (1987) suggests that the inspiration for non-numeric methods derives partly from a prejudice against numerical forms of computation in artificial intelligence that originally eschewed "number-crunching"

[7] An early celebrated expert system for diagnosing microbial infections.

and dealt in symbolic computation. This prejudice he suggests is waning, while interest in probability ideas and numerical uncertainty management is growing. As will become apparent, this thesis is concerned with determining the relative likelihood of competing hypotheses: this requires real numbers to be associated with uncertainty. Consequently the main focus of attention in this survey will be on numerical methods, and non-numeric schemes will not be considered further.

The numerical methods reviewed here are the historically-important Bayesian theory, the Dempster-Shafer theory of evidence, and an important alternative theory based on fuzzy logic. Of necessity the reviewed methods are a tiny subset of the entire panoply of available numerical schemes. Shafer and Pearl (1990) have produced a collection of important papers relating to this field. A comprehensive overview of the techniques has been given by Grzymala-Busse (1991).

3.4 Bayesian Decision Methods

3.4.1 Simplified Bayesian formulation

Probability theory offers a sound logical model for dealing with uncertainty: indeed some assert that it is the only satisfactory description and that alternative methods are unnecessary (Lindley, 1987; Cheeseman, 1985). The classical or *a priori* definition of probability is given by the proportion of cases in which a given event occurs, so that for example the probability of obtaining a one from throwing a fair die is 1/6. However, other interpretations have emerged since the foundation of the theory. In the objectivist interpretation an experimental or *a posteriori* approach is adopted in which the probability of an event occurring is derived from observing the frequency of its occurrence out of a large number of possible outcomes. For the subjectivist interpretation, from which Bayesian theory is derived, a probability is a number in the range zero to one (inclusive) which expresses an individual's confidence in the truth of a proposition.

Subjective probability is a personal belief that will depend on information currently available. Furthermore, two different individuals may hold different degrees of confidence in the truth of a particular proposition (Ng and Abramson, 1990). The key difference between the two interpretations is that while objective probabilities are assigned to repeatable events, subjective probabilities are employed in situations where events are not necessarily reproducible. In this case empirical or frequency data may not exist and the expert's opinion on the likelihood of an event occurring must be sought. An example of such an event may be that of finding oil at a new site, or a stock-market "crash" occurring next year. Such events, for which empirical data may not exist or which is too expensive to collect, are nevertheless of interest to decision makers and expert system builders. Where probability theory is used to represent expert knowledge, as in expert systems, subjective belief is the appropriate interpretation[2].

A personal belief is conditioned on the information available to the individual, i.e. the belief in a proposition given some background knowledge. If the assessor has prior information s, his belief in the event A given the existence of s may be expressed as $P(A|s)$. On observing new evidence B, he may revise his belief in A to $P(A|B,s)$, where B,s denotes the conjunction of event B and background evidence s. Bayesian theory provides a consistent framework for updating beliefs in the light of new evidence. The updating mechanism is Bayes' rule, which written in terms of hypothesis H, and evidence e is (Duda *et al.*, 1976):

$$P(H|e) = \frac{P(e|H)P(H)}{P(e)} \tag{3.1}$$

where the *a priori* probability $P(H)$ for the hypothesis H is modified by the occurrence of e to produce the *a posteriori* conditional probability $P(H|e)$. This is the likelihood of hypothesis H based on some evidence e, or the degree of belief.

To understand this notion of likelihood, it is helpful to express Bayes' rule in the **odds-likelihood** form. Odds are used for conveniently interpreting likelihoods as expressed by conditional probabilities. For example, if the likelihood of an event is 95%, then the odds on the event happening are 0.95 / (1 - 0.95) = 19 to 1. Some people find odds more meaningful than probabilities, although the same information is conveyed (Giarratano and Riley, 1989). Writing the above rule in terms of $\neg H$, the negation of H, and defining the prior odds on H as

$$O(H) = P(H)/P(\neg H) \tag{3.2}$$

the posterior odds on H to be

$$O(H|e) = P(H|e)/P(\neg H|e) \tag{3.3}$$

---

[2] See however Sucar *et al.* (1993) who demonstrate the use of objective probabilities in expert systems.

and the likelihood ratio as

$$L_s = P(e \mid H) / P(e \mid \neg H) \tag{3.4}$$

Bayes' Rule becomes:

$$O(H \mid e) = L_s(e \mid H)\, O(H) \tag{3.5}$$

The prior odds $O(H)$ measures the prospective support for H, without any evidence (or the evidence of "background knowledge" alone), while L_s is the retrospective support for H given the observation of e (Pearl, 1987). The likelihood ratio is also known as the *likelihood of sufficiency*, so called because if $L_s = \infty$, $P(e|H) = 1$ and $P(e|\neg H) = 0$, and so e is sufficient for concluding that H is true (Johnson and Keravnou, 1985). Its value must be supplied by a human expert, or derived from information so provided. It is also possible that the expert may wish to express a subjective opinion of the effect on H given the absence of e. Here the odds-likelihood form is expressed using the likelihood of necessity L_n, defined analogously to L_s but in terms of $\neg e$.

For N pieces of evidence $e^1, \ldots, e^N$ the combined belief in H would be expressed as

$$O(H \mid e^1, e^2, \ldots, e^N) = L_s(e^1, \ldots, e^N \mid H)\, O(H) \tag{3.6}$$

(Pearl, 1987). This equation is problematic in that it requires knowing the conditional probabilities of the mutual occurrence of each set of evidence for hypothesis H and its negation $\neg H$. It is rendered tractable by a fundamental assumption in Bayesian theory: that of conditional independence. If $P(e^1|e^2,H) = P(e^1|H)$ then e^1 and e^2 are said to be conditionally independent given H. This can be written as $P(e^1,e^2|H) = P(e^1|H) \times P(e^2|H)$, so that for each piece of evidence

$$P(e^1, e^2, \ldots, e^N \mid H) = \prod_{k=1}^{N} P(e^k \mid H). \tag{3.7}$$

If the evidence is also conditionally independent under $\neg H$ (the negation of H)

$$P(e^1, e^2, \ldots, e^N \mid \neg H) = \prod_{k=1}^{N} P(e^k \mid \neg H) \tag{3.8}$$

so we can write

$$O(H \mid e^1, e^2, \ldots, e^N) = O(H) \prod_{k=1}^{N} L_s(e^k \mid H) \tag{3.9}$$

where $L_s(e^k|H)$ is the likelihood ratio for each set of evidence. This formula also provides for incremental updating as evidence e is presented: as $O(H)$ becomes $O(H|e)$, the posterior odds become the prior odds for the next iteration.

The requirement for conditional independence under $\neg H$ with H is demanding and may not be reasonable if $\neg H$ entails other states of the world. Pearl (1987) gives an example in which evidence e from a set of alarm sensors supports the hypothesis H that a house has been burgled, so that $\neg H$ corresponds to the state of no burglary. The evidence from the sensors is conditionally independent under both H and $\neg H$ if they are affected by factors solely concerned with their function of detecting intrusion, and not by external factors such as power failure, malfunction, or some other catastrophic event (e.g. an explosion). However, a more reasonable assumption would be that the sensors are indeed affected by such factors. By accounting for these factors in a refined space of hypotheses, evidence from the detectors may be properly regarded as conditionally independent.

Thus, rather than two hypotheses, the domain of interest may often be more appropriately modelled by a set of n hypotheses H_i that are mutually exclusive and exhaustive, i.e.

$$P(H_i, H_j) = 0, \quad i \neq j \tag{3.10}$$

and

$$\sum_{i=1}^{n} P(H_i) = 1 \tag{3.11}$$

After observing N sets of evidence that are conditionally independent with respect to each H_i, the overall belief in the ith hypothesis is (Bonissone, 1987)

$$P(H_i \mid e^1, \ldots, e^N) = \frac{P(H_i) \cdot \prod_{k=1}^{N} P(e^k \mid H_i)}{\sum_{i=1}^{n} P(e^1, \ldots, e^N \mid H_i) P(H_i)} \tag{3.12}$$

Thus the scheme requires N x n conditional probabilities and n - 1 prior probabilities to compute the overall probability distribution over the hypotheses. If no prior knowledge is available, the prior probability for each of the n hypothesis is $1/n$ by the Principle of Indifference.

In practise the denominator of the above equation may be determined from the fact that the left-hand side sums to unity over i since the hypotheses are exhaustive, and therefore may be regarded as a normalising constant (Pearl, 1987). Thus, we can rewrite this as

$$P(H_i \mid e^1,\ldots,e^N) = \text{constant} \times P(H_i) \cdot \prod_{k=1}^{N} P(e^k \mid H_i) \qquad \textbf{(3.13)}$$

which is similar to the odds-likelihood formulation of Bayes' rule for a single hypothesis.

The multi-hypothesis case can also be expressed in the odds-likelihood formulation using the **Modified Bayesian Rule** (Bonissone, 1987). For this each piece of evidence e^k must be conditionally independent under $\neg H_i$ and H_i, i.e.

$$P(e^1,e^2,\ldots,e^N \mid \neg H_i) = \prod_{k=1}^{N} P(e^k \mid \neg H_i) \qquad \textbf{(3.14)}$$

However Pearl (1982) points out that conditional independence with respect to the negations of the hypotheses is normally violated (i.e. $P(e^1,e^2 \mid \neg H) \neq P(e^1 \mid \neg H)\, P(e^2 \mid \neg H)$). Furthermore, the elicitation of the $P(e^k \mid \neg H)$ is not a natural mental task. If $\neg H$ is subsumed within an enlarged multi-hypothesis space, Bayes' rule can be given solely in terms of the H_i rather than their negations, as in Pearl's formulation. In other words, the multi-hypothesis model can be developed solely in terms of the H_i rather than $\neg H_i$, thereby making any references to odds unnecessary.

3.4.2 Rule-based and graphical representations

3.4.2.1 Problems with the simplified formulation

Both the single hypothesis and multi-hypothesis model are part of the simplified Bayesian formulation, in that they are based on two simplifying assumptions of (1) mutual exclusivity and exhaustiveness and (2) conditional independence of the evidence. The model is well suited to diagnostic tasks in which the hypotheses represent some proposition whose likelihood we wish to determine (e.g. bacteriological infection, existence of oil below ground, quality of river water at a site) which are supported or refuted by evidence as various symptoms or observations (e.g. morphology of infecting organism, seismic survey results, presence or absence of benthic taxa). **Figure 3.1** shows a *belief network* representing the simplified Bayesian formulation in which a single hypothesis node H representing n mutually exclusive and exhaustive hypotheses that lead to a set of observations or evidence $e^1 ... e^N$. The absence of arcs between the items of evidence is a consequence of the assumption of conditional independence, i.e. each observation is independent of another given any one of the hypotheses H_i.

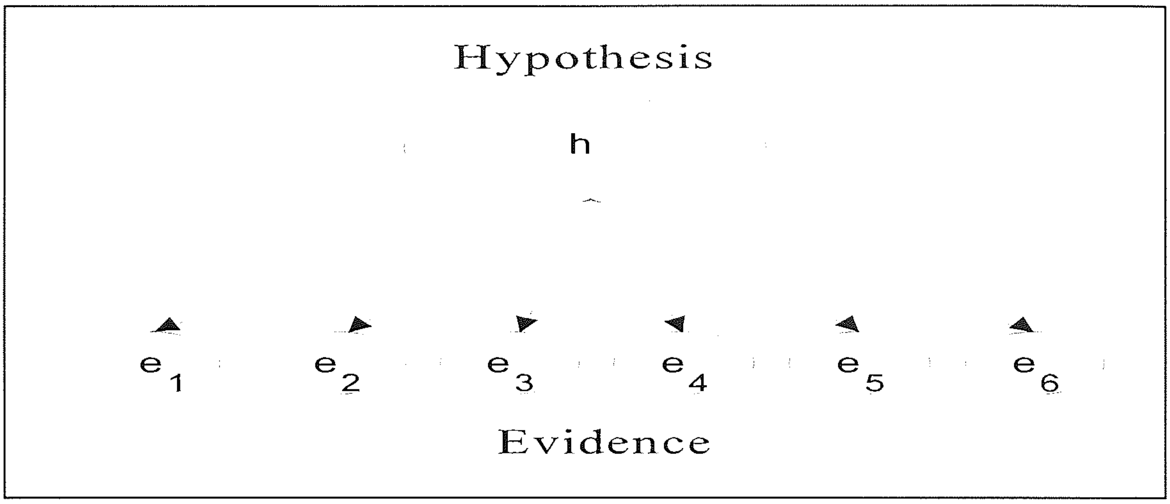

Figure 3.1 Belief network for simplified Bayesian formulation

However the simplified Bayesian formulation has been criticised because of its two key assumptions, which are regarded by some to be too restrictive. Szolovits and Pauker (1978) in their examination of medical diagnostic reasoning consider that the hypotheses in most clinical situations are neither exhaustive nor mutually exclusive. Performing Bayesian calculations in a hypothesis space that is not exhaustive lead to incorrect posterior probabilities, while insisting on mutual exclusivity requires the creation of hypotheses corresponding to every possible combination of diseases. In fact the authors maintain that clinical decision making in the medical domain requires a judicious mixture of probabilistic and *categorical* reasoning: the use of rules that lead to unambiguous and explicit judgements.

3.4.2.2 Rule-based systems

Rule-based representations of knowledge are the most familiar and popular structures for building expert systems for several reasons. Rules are normally easy to write, provide a highly visible representation of the expert's knowledge, and can be added incrementally to the knowledge base. Hierarchical relationships between rules can be encoded, facilitating explanations of conclusions by "tracing back" rules fired by other rules. Production rules are moreover considered to mirror the human cognitive process (Shadbolt, 1989).

Rules encoded as unqualified IF...THEN statements (i.e. if the antecedent holds, the consequent is true, with absolute certainty) do not however reflect the inherent uncertainty in human knowledge. To cope with uncertain reasoning, some early rule-based expert systems such as PROSPECTOR and MYCIN extended the deterministic approach to incorporate numerical degrees of truth between a rule's premise and its conclusion, effectively to weaken the connection between its antecedent and consequent (Bonissone, 1987). A typical rule from MYCIN, developed for diagnosing bacterial infections, illustrates this approach (Buchanan and Shortliffe, 1985):

IF:	1)	The stain of the organism is gram positive, and
	2)	The morphology of the organism is coccus, and
	3)	The growth conformation of the organism is chains
THEN:		There is suggestive evidence (.7) that the identity of the organism is streptococcus

The strength of the evidence supporting a particular hypothesis is weighted here by a *certainty factor*. In this, as with all numerical schemes for handling uncertainty, the individual beliefs are combined to decide their joint effect. Rather than arriving at a categorical decision regarding a hypothesis in an exact reasoning scheme, several hypotheses are supported to differing degrees.

3.4.2.3 Problems with rule-based representations

A consequence of reasoning with certain rules is that once the truth of a proposition has been asserted it cannot be changed by other facts, i.e. the logic with which the system reasons is monotonic. This property facilitates the modularity of the system, by which means rules can be added incrementally to the knowledge base. However according to Shafer (1987) "the introduction of probabilities into production rules does not square well with the modularity we want these rules to have", so that rule-based representations have problems in dealing adequately with uncertainty.

Consider for example the *inference network* shown in **Figure 3.2** in which observations of evidence e_i are used to measure support for the hypotheses H_i. Here the intermediate evidence (e.g. e_1 , e_2 , e_3) can be thought of as hypotheses to be supported by underlying evidence (e_4, e_5). As this underlying evidence is accumulated, the appropriate production rules are fired and posterior probabilities are calculated, propagating through the network via the inference links towards the sought hypotheses H_i.

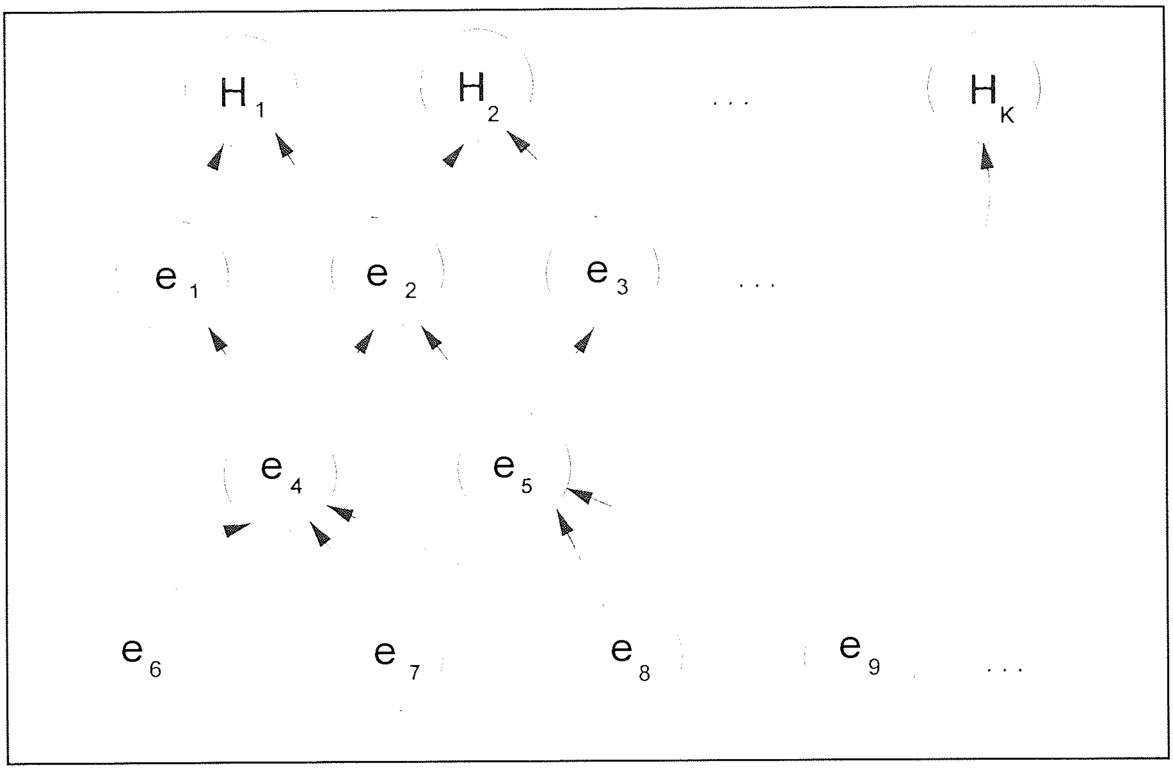

Figure 3.2 Rule-based inference network (after Duda *et al.*, 1976)

Problems arise both in eliciting the conditional probabilities associated with these rules and propagating them through the network. In the first case the expert may be able to supply rules for probabilities $P(e_5|e_8)$ or $P(e_5|e_9)$ but be unable to do so for $P(e_5|e_8 \cap e_9)$. In other words, a joint-probability distribution may not be available for all the variables, leading to inconsistent probabilities. On the other hand the conditional probabilities may effectively constrain the expert's knowledge. The joint probability[3] of e_8 and e_9 is fixed by the prior probabilities $P(e_8)$ and $P(e_9)$, since

$$P(e_8 \cap e_9) = P(e_8)P(e_9)[P(e_5|e_8)P(e_5|e_9) + P(\neg e_5|e_8)P(\neg e_5|e_9)] \tag{3.15}$$

This means that the expert is not free to supply the joint probability if this is known, since it is predetermined by the prior probabilities. A similar problem arises in the specification of L_s and L_n, the likelihoods of sufficiency and necessity discussed previously. In PROSPECTOR, the experts were asked to supply these values rather than probabilities. However an expert is not able to estimate L_s and L_n freely, whatever his own judgements

[3] A joint probability distribution is one in which the probabilities of a set of random variables are defined for all values of the variables.

may dictate, since these quantities are dependent:

$$L_n = \frac{1 - L_s P(e|\neg H)}{1 - P(e|\neg H)} \tag{3.16}$$

This relation follows from the definitions of L_s and L_n. The problem here is that while the observation of evidence e may be considered to support hypothesis H, the absence of e may not necessarily be viewed by the expert as disconfirming H. In other words, while L_s may be > 1, in the expert's view L_n may be equal to 1, in which case the absence of e makes no impact on his judgement regarding the likelihood of H. Such an opinion however would be inconsistent with the conditional probabilities as expressed by equation **(3.16)**. If the evidence itself is uncertain, this inconsistency leads to further problems when propagating probabilities, which can be significant for long chains of inference. To overcome this difficulty the PROSPECTOR system used *ad hoc* adjustments to the propagation formula.

A more important problem lies in the way probabilistic knowledge is represented. In diagnostic systems, rules are generally encoded in the direction of evidence to the hypotheses: if e then H (with probability x). With this structure reasoning predictively is difficult: if H obtains, what is the likely effect (i.e. what findings would we expect?). In Bayesian reasoning expert knowledge is often elicited to express this causality, but combined in such as way as to effect a diagnosis. Thus the conditional probabilities $P(e|H)$ linking cause (the hypothesis) to effect (the evidence) are combined via Bayes' rule to produce diagnostic support for H given all the available evidence.

Problems also arise due to the effects of multiple causes. **Figure 3.3** shows an example from Pearl (1988). If a burglar alarm sounds, there is a high probability that it was caused by an attempted burglary. However, the alarm may also be triggered by an earthquake. Independent evidence for such an event, such as a report in the newspapers or on radio should reduce support for the burglary hypothesis by "explaining away" the alarm sound. While special rules can be written to mirror this effect, they effectively encode inference procedures rather than domain knowledge. In contrast to this, probabilistic methods can represent this knowledge more naturally, and allow for both diagnostic and predictive inference.

Figure 3.3 Independent evidence of an earthquake (the radio report) "explains away" the alarm, thereby reducing support for the burglary (from Pearl (1988)).

An inference network as depicted in **Figure 3.2** is determined by the underlying rule set, rather than any causal relationship between assertions.[4] As argued above it is usual to gather evidence at the root of the network as support for intermediate hypotheses, which in turn lend support for the sought top-level hypotheses. Thus, probability propagation is effectively one-way. Attempts to incorporate causal relationships characterised by uncertainty into rule-based systems leads to overly-complex inference networks that obscure the relationships between assertions. Neapolitan (1990) argues that the modelling of complex probabilistic relationships, a task that is apparently quite natural to humans, is neither feasible nor reasonable using a rule-based approach.

3.4.2.4 Bayesian graphical methods

The limitations of both the simple Bayesian formulation and rule-based representations discussed above have led to recent interest in graphical methods. In decision analysis, an increasingly popular method of modelling problems characterised by uncertainty is via the use of *influence diagrams*, graphical representations of uncertain and decision variables that explicitly depict probabilistic dependence and independence. These diagrams are networks of nodes, representing probabilistic or deterministic variables and directed arcs that represent the relationships between them. *Chance nodes*, depicted as circles or ovals on influence diagrams, represent states of the world that are uncertain. They correspond to propositions (random variables) in the problem.

---

[4] The term assertion is used as a generic term for both evidence and hypotheses in an inference network.

While these *directed, acyclic graphs* may be formally described and evaluated, their popularity stems largely from the clarity with which they can depict and elicit qualitative knowledge concerned with decision problems (Shachter, 1986). Where the diagram involves only random variables (chance nodes), it is called a Bayesian or *belief network* (Pearl, 1986a) or causal or independence network (Neapolitan, 1990). The term *causal* is useful in distinguishing this technique from inference networks. In belief or causal networks, the arcs link cause and effect, whereas with inference networks the arcs link evidence to hypothesis.

Belief networks are a method of probabilistic reasoning that have recently been the focus of a considerable research effort: see Charniak (1991) for a comprehensive list of references. An example of a single-layer network was given in **Figure 3.1**, corresponding to the simplified Bayesian formulation discussed above. A richer knowledge scheme is represented by **Figure 3.4** that illustrates the different kinds of independence in a network. The source variables *a* and *c* have no predecessors and are therefore marginally (i.e. unconditionally) independent. Variables *b* and *d* are conditionally independent given their common predecessor *c*, while *e* is conditionally independent of *a* and *c* given *b* and *d*, its

Figure 3.4 Independence relations in a belief network. Given knowledge of *c*, variables *b* and *d* are conditionally independent (from Henrion *et al.* (1991))

immediate predecessors. The independencies depicted in the belief network also illustrate localised effects between variables: the effects of one variable on a distant variable can only propagate along the influence arcs.

3.4.2.5 Evaluating belief networks

Probabilities assigned to the random variables in a belief network form a joint-probability distribution uniquely defined as the product of the individual distributions for each random variable, themselves equal to the variable's conditional probability given its predecessors. Thus, the joint distribution for the network of **Figure 3.4** can be written as:

$$P(a,b,c,d,e) = P(e \mid b,d)\, P(b \mid a,c)\, P(d \mid c)\, P(c)\, P(a) \tag{3.17}$$

Once a belief network is constructed, it can be used to interpret input data, on the basis of equation (**3.17**). A set of variables corresponding to the data is instantiated, and the effect propagated through the network to evaluate the new probabilities of a set of variables designated as the hypotheses. Evaluation of a network therefore consists of calculating every node's conditional probability from the evidence available.

Unfortunately this evaluation becomes infeasible for networks of more than about 10 nodes, since the computation is exponential in the number of variables and is said to be NP-hard.[5] If however the network is singly-connected, with no more than one path between any pair of nodes, then exact solutions are possible. Multiply-connected networks, which obtain in most realistic applications, can be solved exactly if transformed into their equivalent singly-connected graphs. Usually however such networks require approximate solutions to the conditional probabilities: the reader is referred to Henrion (1988, 1990) for a detailed description of these.

Advocates for belief networks in uncertain reasoning cite several benefits. On the practical side large-scale belief networks have been recently constructed for knowledge-intensive applications with some success: examples include the commercial INTELLIPATH system for the diagnosis of lymph node diseases and MUNIN for diagnosis of neuromuscular disorders, and many more applications are envisaged (Horvitz *et al.*, 1988). Theoretically the approaches of decision analysis and probabilistic reasoning can be integrated via the use of influence diagrams to express both uncertain knowledge and the utility of decisions to be made under uncertainty. Furthermore, the graphical nature of this knowledge representation promotes clarity and comprehension of the decision problem. Belief networks and influence diagrams explicitly depict dependence and independence between variables. Reasoning in networks may be in any direction: causal (or predictive), diagnostic, or intercausal, overcoming a particular drawback of rule-based inference that is usually from evidence to hypothesis, and requires special rules to deal with inter-causal effects.

---

[5] Non-deterministic, Polynomial time.

Drawbacks include the problems of achieving exact solutions, and the approximations required for representing multiply-connected graphs as singly-connected (Shafer, 1987). Charniak (1991) acknowledges a single satisfactory algorithm for all network types may not exist. Moreover, the knowledge acquisition task for building belief networks should not be underestimated. The independence assumptions in Bayesian networks follow from a causal interpretation, i.e. the arcs linking the nodes are from cause to effect. Thus, constructing a network requires the expert to be familiar with the cause and effect relationships among all the variables, and to be able to identify the direct cause of each variable (Neapolitan, 1990).

3.5 Dempster-Shafer Calculus

3.5.1 Introduction to Dempster-Shafer theory

Several alternative formalisms for uncertainty reasoning, such as certainty-factors, have been devised in response to perceived problems with Bayesian theory. One such alternative that has received substantial attention as a viable and practical non-Bayesian reasoning method is Dempster-Shafer or Belief theory, a system for manipulating degrees of belief that do not require assumptions about the prior probabilities of hypotheses under consideration (Tanimoto, 1990). If we have very little evidence for a proposition, we can accord little or no belief to the proposition and its negation. The calculus is therefore more general than the Bayesian approach.

Gordon and Shortliffe (1985a) have commended the method for its ability to closely model diagnostic reasoning and to distinguish between uncertainty, or lack of knowledge, and indifference. The theory moreover has a firm mathematical foundation, emerging from the study of multi valued mappings by Dempster (1967) and developed into a theory of evidence by Shafer (1976). This contrasts with the essentially *ad hoc* nature of certainty factors (Gordon and Shortliffe (1985a), Giarratano and Riley (1990)). The theory has found practical uses in a variety of applications, including the analysis of sleep disorders (Principe *et al.*, 1989), aircraft target identification by multiple sensors (Bogler, 1987), computer vision systems (Andress and Kak, 1987), surface cover classification in remote sensing (Pendle and Franklin, 1992), foreign exchange-rate forecasting (Ip and Wong, 1991) and forecasting and marketing management (Cortes-Rello and Golshani, 1990).

In the following discussion the basic ideas of the theory are explained in terms of a model of river water quality classification using biological data. Five classes B1a, B1b,B2,B3,B4 are adopted, parallelling the chemical classification currently used by the National Rivers Authority in England and Wales and ranging from unpolluted waters (B1a) to grossly polluted (B4).[6] As with probability theory, Dempster-Shafer reasoning assumes a fixed set of mutually exclusive and exhaustive hypotheses (corresponding to the sample space S) known as the **frame of discernment** Θ containing the exhaustive set of propositions in the domain of interest. In our case this is the set of biological water classes: thus $\Theta = \{B1a,B1b,B2,B3,B4\}$.

Each hypothesis in Θ corresponds to a one-element subset (called a *singleton*). However in D-S reasoning the term *hypothesis* is normally used in an enlarged sense, since one can allocate belief to all the possible subsets of Θ (the "power set") totalling $2^{|\Theta|}$ in number, in contrast to the $|\Theta|$ hypotheses of the Bayesian sample space. Thus for a frame

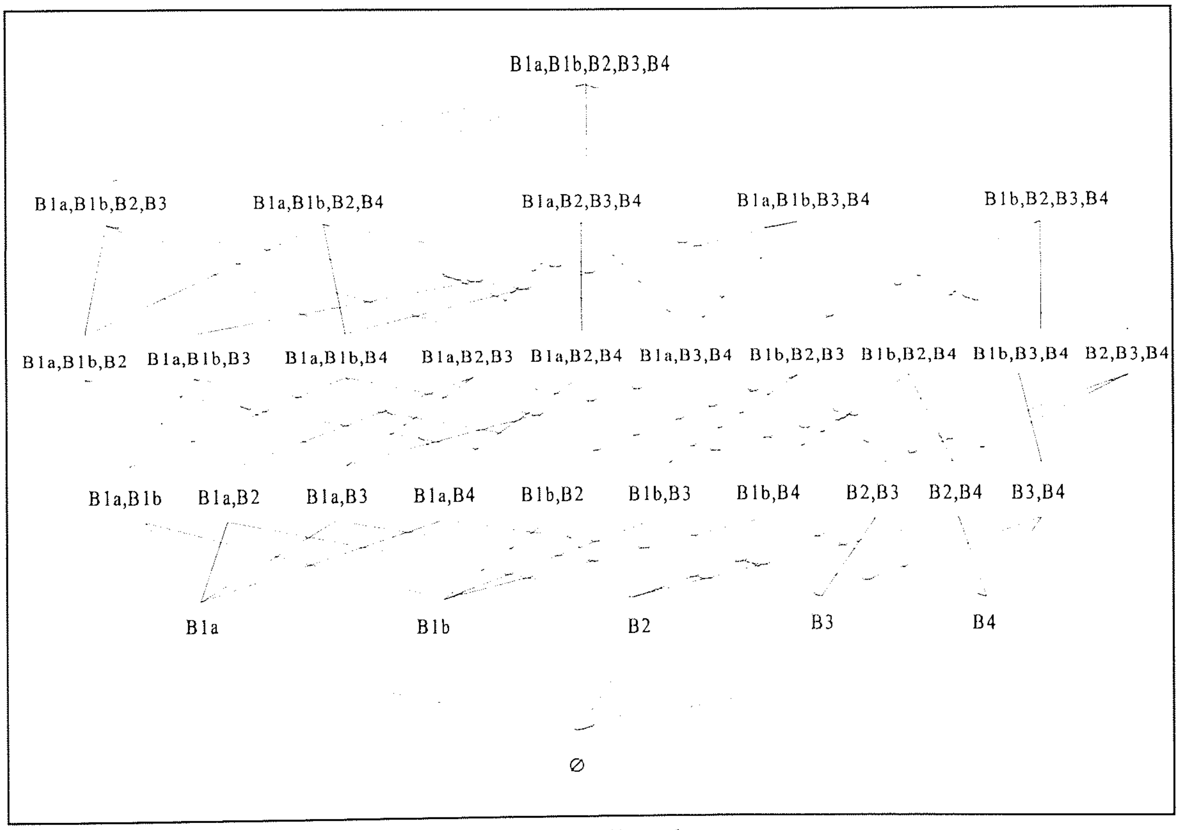

Figure 3.5 The subsets of the set of water quality classes

with five elements, there are $2^5 = 32$ subsets, including the null set $\varnothing$. **Figure 3.5** shows the power set for the frame of discernment formed from the set of biological water quality classes $\{B1a,B1b,B2,B3,B4\}$.

The subsets can be viewed as corresponding to various propositions of diagnostic

[6] See discussion in Chapter 2 on discrete biological classification.

interest in the power set. As an example, the proposition 'Poor Quality water' could be represented by the two-element subset {B3,B4}, while its complement ¬{B3,B4} corresponds to the three-element subset {B1a,B1b,B2} representing the proposition 'not Poor Quality water' or perhaps 'Good to Fair Quality water'. In Dempster-Shafer reasoning evidence against an hypothesis is equivalent to evidence supporting the complement of that hypothesis, avoiding the use of negative numbers as in the certainty-factor model (Gordon and Shortliffe, 1985).

3.5.2 Basic probability assignments

A number m in the range [0,1] known as the *basic probability assignment* (*bpa*) is used to represent the degree to which some evidence supports the various propositions. Formally, a basic probability assignment is a function that maps each element of the power set into a real number in [0,1], i.e.

$$m: 2^{\Theta} \rightarrow [0,1] \qquad \textbf{(3.18)}$$

Further, since the null set $\varnothing$ is the hypothesis known to be false, $m(\varnothing) = 0$, and the total mass across the power set assigned by an item of evidence must sum to unity:

$$\sum_{A \in 2^{\Theta}} m(A) = 1 \qquad \textbf{(3.19)}$$

Note that only those propositions for which supporting evidence is available need be assigned a probability mass: if further evidence is unavailable the remaining support need not be committed to any particular subset of the frame: it can instead be assigned to the environment Θ. As an example, consider that one piece of evidence from a sampled site results from observing that the freshwater shrimp *Gammarus pulex* is present. Allowing for a certain degree of uncommitted belief, the basic probability assignment suggested by this evidence may be as follows: $m(\{B1a\}) = 0.2$, $m(\{B1b\}) = 0.5$, $m(\{B2\}) = 0.1$. The uncommitted belief $m(\Theta)$ is then equal to 0.2 by definition of a *bpa*. Thus uncertainty can be explicitly represented by assigning uncommitted belief to Θ, rather than to any proper subset of it. If evidence favours a single subset, say {B1b,B2}, remaining belief can be assigned to Θ rather than to the complement of this subset, as would be required in the Bayesian. Dempster-Shafer's treatment of ignorance is therefore markedly different from that in probability theory.

Evidence is thus provided by a reference set of benthic organisms that may be viewed as sensors, able to report their current state and accordingly assign a measure of support to the relevant propositions. The inference task then becomes one of integrating evidence from these disparate, perhaps conflicting sources to produce a report of overall support for the biological classes. This problem of fusing evidential data from multiple knowledge sources using Dempster-Shafer reasoning has been examined by Garvey *et al.* (1981) and notably Bogler (1987) in an application of aircraft target identification.

3.5.3 Belief functions

One consequence of the flexibility of belief assignment in Dempster-Shafer reasoning is that different levels of precision, or granularity, of belief parameters may be employed. The basic probability number $m(A)$ represents the belief committed exactly to the subset A, which cannot be subdivided further among any of its subsets. The *total* belief in A, denoted $Bel(A)$, is measured by the *bpa* assigned to A and all its subsets, i.e.

$$Bel(A) = \sum_{B \subseteq A} m(B) \qquad \textbf{(3.20)}$$

From the above $m(A) = Bel(A)$ for all singletons. The definition is reasonable in that belief in a set enhances support for its superset. While $Bel(A)$ describes the belief that A is definitely true, the plausibility of A, $Pls(A)$ represents the belief that A could be true, i.e. the degree of belief that would exist if all remaining uncertainty was removed. Thus

$$Pls(A) = 1 - Bel(\neg A) \qquad \textbf{(3.21)}$$

In Bayesian reasoning $Pls(A) = Bel(A)$, but in Dempster-Shafer theory it is usual to express belief in a proposition in terms of the belief interval $[Bel(A), Pls(A)]$ which is usually of non-zero width. The width of this interval can be regarded as the amount of uncertainty in a proposition, the degree to which we are unsure whether or not it is true. It is therefore belief committed by the evidence to neither the hypothesis nor its negation. For some subset A, this probability mass is associated with all those supersets of A which intersect it. If more specific evidence became available, the mass currently with a superset could be committed to A, thereby reducing the width of the belief interval and therefore the degree of uncertainty.

3.5.4 Combination of belief functions

Support for the various hypotheses arises from observing multiple items of independent evidence. This support must then be combined in a way that represents the aggregation of this evidence. In D-S reasoning this combination takes place by Dempster's rule (Shafer, 1976). For two sets of evidence with *bpa*'s m_1 and m_2 respectively, Dempster's rule computes a new *bpa* representing their combined effect and called the *orthogonal sum* of m_1 and m_2. For some proposition Z of Θ, $m_1 \oplus m_2(Z)$ is the sum of all products of $m_1(X)$ and $m_2(Y)$, where X and Y are subsets of Θ whose intersection is Z. Formally,

$$m_1 \oplus m_2(Z) = \sum_{X \cap Y = Z} m_1(X)\, m_2(Y) \,/\, (1 - \kappa) \qquad \textbf{(3.22)}$$

where

$$\kappa = \sum_{X \cap Y = \emptyset} m_1(X)\, m_2(Y) \qquad \textbf{(3.23)}$$

κ corresponds to the degree of evidential conflict between the two sources. If the sources do not conflict in any way, κ equals zero. If however the sources do not have any sets in common, the evidence is totally contradictory and κ equals unity, so that here the orthogonal sum given by **(3.22)** is not defined. Dempster's rule promotes consensus by only assigning belief to intersecting sets, which represent common elements of evidence. Since the rule is commutative, the evidence may be combined in any order. Gordon and Shortliffe (1985b) note the importance of this property in diagnostic reasoning.

A simple example from Boyd *et al.* (1993) will illustrate Dempster's rule. Suppose that the presence in a sample of the mayfly *Baetis rhodani* supports the proposition {B1a,B1b,B2} to degree 0.6 (m_1) while the presence of Lumbriculidae supports {B1b, B2,B3} to degree 0.7 (m_2). What is the net effect of these two items of evidence? The computation of the orthogonal sum is shown in **Table 3.1**.

Table 3.1 Illustration of Dempster's rule for two basic probability assignments

	m_2 ({B1b,B2,B3}) (0.7)	m_2 (Θ) (0.3)
m_1 ({B1a,B1b,B2}) (0.6)	{B1b,B2} (0.42)	{B1a,B1b,B2} (0.18)
m_1 (Θ) (0.4)	{B1b,B2,B3} (0.28)	Θ (0.12)

The first row of the table shows the first *bpa* m_1 in which a belief of 0.4 is left unassigned to the environment Θ. The first column shows the second *bpa* m_2. Set intersections are shown in the table alongside the numeric mass product. For instance the set intersection of {B1a,B1b,B2} and {B1b,B2,B3} is {B1b,B2}, while the basic probability number assigned to this new proposition is 0.6 x 0.7 = 0.42. The orthogonal sum is thus: $m_1 \oplus m_2$({B1b,B2}) = 0.42, $m_1 \oplus m_2$({B1b,B2,B3}) = 0.28, $m_1 \oplus m_2$({B1a,B1b,B2}) = 0.18, $m_1 \oplus m_2$ (Θ) = 0.12. The belief functions can then be recovered from the *bpa*s. For example, *Bel*({B1b,B2,B3}) = $m_1 \oplus m_2$({B1b,B2,B3}) + $m_1 \oplus m_2$({B1b,B2}) = 0.42 + 0.28 = 0.7.

3.5.5 Evidential conflict

In the above example, no conflict occurs between the two evidential sources. Consider now the effect of further evidence arising from the observation that sludge-worms are present in some numbers, suggesting that the sampled site may be affected by organic pollution. The *bpa* induced by this evidence may be as follows: m_3({B3,B4}) = 0.8, $m_3(\Theta)$ = 0.2. Here, the probability mass is focused onto more specific hypotheses (the poorer quality waters) with a higher degree of certainty. The uncommitted evidence is likewise reduced, suggesting more confidence in the data quality of this source. However there is a degree of conflict between this and the previous *bpa*, as can be seen in the intersection matrix shown in **Table 3.2**. It can be seen from the table that in two cases the combining subsets have no hypotheses in common, i.e. their intersection is the null set. The probability mass associated with the null set is thus 0.336 + 0.144 = 0.48. However, by definition of a *bpa* the mass assigned to the null set must be zero, since this is the false hypothesis.

Table 3.2 Illustration of Dempster's rule for conflicting evidence

	m_3 ({B3,B4}) (0.8)	m_3 (Θ) (0.2)
$m_1 \oplus m_2$({B1b,B2}) (0.42)	$\varnothing$ (0.336)	{B1b,B2} (0.084)
$m_1 \oplus m_2$({B1b,B2,B3}) (0.28)	$\varnothing$ (0.144)	{B1a,B1b,B2} (0.036)
$m_1 \oplus m_2$({B1a,B1b,B2}) (0.18)	{B3} (0.224)	{B1b,B2,B3} (0.056)
$m_1 \oplus m_2$ (Θ) (0.12)	{B3,B4} (0.096)	Θ (0.024)

To deal with this apparent anomaly $m_1 \oplus m_2 \oplus m$ {$\varnothing$) is set to zero and the remaining probability numbers are <u>normalised</u> by dividing by the factor (1 - 0.48) = 0.52, as shown for the following two subsets:

$$m_1 \oplus m_2 \oplus m_3 \ (\{B3\}) \quad = 0.224/0.52 = 0.43$$
$$m_1 \oplus m_2 \oplus m_3 \ (\{B3, B4\}) = 0.096/0.52 = 0.18$$

The resulting *bpa* across the subsets sums to unity as required.

3.5.5.1 Problems with Dempster's rule

The process of normalisation in Dempster's rule redistributes the conflicting probability mass across the "consenting" subsets, in a way that maintains the relative degree of support for each. Gordon and Shortliffe (1985) note however that this process is accepted by way of convention rather than by theory, and according to Zadeh can lead to inconsistent behaviour if the evidence is highly conflicting. As an example (adapted from Zadeh (1984)) consider just three hypotheses regarding the biological quality of river water: (a) very good, (b) medium and (c) very poor. Two experts (or sets of evidence) may make the following assessments:

$$m_1(a) = 0.99 \quad m_1(b) = 0.01 \quad m_1(c) = 0$$
$$m_2(a) = 0 \quad m_2(b) = 0.01 \quad m_2(c) = 0.99$$

The first expert is almost certain that the water is of high quality, that medium quality is highly unlikely, and that the proposition of the water being of very poor quality can be rejected. The second expert's opinion is exactly the opposite. Combining these two assessments using Dempster's rule results in the following. Since:

$$\kappa = (0.99 \times 0.01) + (0.99 \times 0.99) + (0.01 \times 0.99) = 0.9999$$

we have

$$m(a) = (0.99 \times 0 \quad) \,/\, (1 - \kappa) = 0$$
$$m(b) = (0.01 \times 0.01) \,/\, (1 - \kappa) = 0.0001/0.0001 = 1$$
$$m(c) = (0 \quad \times 0.99) \,/\, (1 - \kappa) = 0$$

This result is counterintuitive. Those hypotheses strongly favoured by one or other of the experts have been ruled out after this combination, while the proposition that both considered being highly unlikely is concluded to be certain beyond doubt.

The problem here arises from the fact that the evidence is both highly contentious and crisp (Zadeh, 1984). Given the high degree of conflict, one should suspect and if necessary reject one or both sets of evidence. Such a decision could be taken by the reasoning system or decision maker by monitoring the value of κ, the degree of evidential conflict in combination. Alternatively, the "crispness" of the evidence may be reduced by discounting, in which part of the *bpa* is removed from focal elements (the subsets within the power set) and reassigned to Θ. Given these difficulties certain workers have considered alternatives or modifications to Dempster's rule, or examined its theoretical justification (e.g. Voorbraak (1991)). Alternatives to Dempster's rule for combining biological evidence are considered in the chapters on the classification experiments.

3.5.6 Representation of belief (Classes of belief function)

One distinguishing feature of Dempster-Shafer reasoning is its flexibility for assigning belief at different levels of granularity, or specificity, across the power set of hypotheses. This flexibility however brings its own problems and is a source of criticism from those who favour Bayesian methods (e.g. Kyburg (1987), Ng and Abramson (1990)). Potentially, the process of assigning belief can be overwhelming: for the $2^{|\Theta|}$ subsets of the power set, there is a total of 2 *exp* $2^{|\Theta|}$ possible basic probability assignments (Kyburg, 1987). In practice, an expert or reasoning system employing Dempster-Shafer theory would need to restrict the number of ways in which probability mass could be distributed by selecting an appropriate class of belief function, an approach adopted by Caselton *et al.* (1988). Once selected, these belief function classes effectively constrain the possible mass assignments.

Several classes of belief function are considered in later chapters dealing with the classification experiments: Bayesian belief, singleton support, simple support and consonant belief. Discussion on their representation and implementation will be delayed until then.

3.6 Possibility Theory

One of the sources of uncertainty is the ambiguity and vagueness of natural language used by human experts. Examples include expressions such as "cold", "tall" or "very probably". The need to cope with this imprecision led to a theory of uncertainty proposed by Zadeh (1978) based on fuzzy logic, itself an extension of fuzzy set theory (Zadeh, 1965). Possibility theory provides a means by which such imprecise knowledge can be represented and as such is another mechanism for reasoning with uncertain evidence. Only a brief outline of possibility theory will be given here as a means of describing its use in evidential reasoning. For introductory readings see for example Graham and Jones (1988), Barron (1993).

In contrast to classical set theory where an object (x) either is or is not a member of a set (A), fuzzy sets allow partial membership. This degree of membership is represented by the membership function $\mu_A(x)$ which takes values in the interval [0,1] inclusive. In classical theory only the two extreme values are allowed: a value of one implies total belief that the element belongs to the set A, a value of zero means that is that is does not. For fuzzy sets, the membership function can be any value in the interval. For example consider the set (X) of earthquake magnitudes X = [4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5] (Wong *et al.*, 1986). If A is the fuzzy subset "a moderate earthquake", it may be subjectively characterised as:

$$A = [\,0/4.5 + 0.2/5.0 + 0.8/5.5 + 0.95/6.0 + 0.8/6.5 + 0.1/7.0 + 0/7.5\,] \tag{3.24}$$

The numerator refers to the membership function of the element in the denominator, so we see that an earthquake of magnitude 6.0 is highly likely to be considered as moderate, magnitudes 5.5 and 6.5 are likely to be considered as moderate, while magnitudes 4.5 and 7.5 would certainly not be considered moderate.

Possibility distributions arise from considering the capability that an element belongs to a set, and therefore are directly related to fuzzy membership functions. So in the above example the element representing earthquake magnitude 4.5 cannot belong to the fuzzy set A, while it is highly possible that 6.5 does. Note that the possibility of an event is very different from its probability: while an earthquake of magnitude 6.0 is highly likely to be considered as moderate ($\mu_A(6.0) = 0.95$), the probability of such an event occurring is to be determined from statistical data. To understand this distinction further it must be remembered that probability theory and related methods such as Dempster-Shafer theory deal with propositions that are "crisp": either true or false. Exactly one proposition in the domain of interest will be true, but before evidence is brought to bear we are unsure which one this is. The propositions themselves are precise: "the river water quality is {B3}", however it is uncertain how true this proposition is, so a number is associated with the proposition to express this uncertainty. In fuzzy set theory however, the propositions are vague: "the earthquake is moderate". Equation **(3.24)** is not an expression of uncertainty (or truth): it is an expression of vagueness. Although an earthquake of magnitude 6.0 is highly likely to be considered moderate, there is no way of finding out for certain whether or not it belongs to the equivalent fuzzy subset, since there is no uncertainty to remove.

Because of this distinction, the class of problems addressed by fuzzy set theory is fundamentally different from those dealt with by probability theory and related methods. In this respect therefore, fuzzy set theory is beyond the scope of this project, which is concerned with reasoning methods applied to propositions that are definitely either true or false. It is discussed in this review because of its importance and because there are overlaps between fuzzy set theory and evidential-reasoning systems. Evidence represented by the fuzzy sets can be combined to infer some conclusion.

3.7 Comparison with Neural Network approach

This project was one of two investigating the application of techniques from Artificial Intelligence to the biological classification of river water quality. The related project using neural networks has been alluded to previously. In this section, a brief review of the differences in the knowledge representation of expert systems and in artificial neural networks is given.

Knowledge-based systems provide a means of reaching a decision by categorical rules or for reasoning with uncertain facts to decide the likelihood of an hypothesis of interest. Knowledge as production rules, probabilistic dependence relations or conditional probabilities is explicit, usually obtained from experts. The classifiers developed for this project are reasoning systems based on this explicit prior knowledge. To process a larger set of input variables the knowledge base would need to be expanded. If a Bayesian network approach was employed, dependence relationships between new nodes and the existing network would need to be made explicit.

For the neural network approach, the net learns by adjusting the weights connecting its processing elements in response to input and required output data. Neural networks exhibit plasticity, in that the net can be retrained even if the number of input elements is reduced or enlarged. In this respect therefore the neural net approach is more adaptable for this type of classification problem. Nevertheless the uncertain reasoning approach is attractive in its explicit representation of quantitative probabilistic knowledge, and for the several mathematically coherent methods available for reaching decisions under uncertainty. Ruck (1995) discusses the possibility of combing the two approaches, which can be seen as complementary. The Bayesian network alternative is considered in Chapter 8.

3.8 Uncertainty Management: Concluding Remarks

This chapter has reviewed several numerical schemes for managing uncertainty, and their role in artificial intelligence. Bayesian decision methods remain the primary numerical approach for representing and manipulating uncertainty. They have a sound mathematical foundation and a wide range of practical applications in expert systems and decision theory. They can be used in simple inference networks with considerable effect although problems arise when rule-based inference networks are extended to incorporate uncertainty. Neapolitan (1990) considers that a mixture of rule-based systems and probabilistic reasoning is infeasible. However it is recognised that the simplified Bayesian scheme can be used successfully in small, well-defined problem domains (Szolovits and Pauker (1978) and Henrion *et al.* (1991)). Causal or Bayesian networks illustrate their power and potential for complex domains, but solutions to such networks are generally approximate.

Dempster-Shafer theory is a generalisation of Bayesian theory that allows considerable flexibility in the representation of uncertain knowledge. In particular it allows ignorance to be explicitly expressed via the mechanism of uncommitted belief. The theory has been criticised on the grounds of its complexity and the lack of guidelines for assigning belief across the power set. As suggested in this chapter however, there are classes of belief functions corresponding to intuitive representations of evidence that constrain the possible assignments that can be made.

An alternative uncertain reasoning paradigm based on fuzzy logic was briefly reviewed. It was noted that the class of problems addressed by fuzzy set theory is different from those dealt with probability theory and related methods. The neural network approach was compared with the knowledge-based approach in the context of the biological classification problem.

The Bayesian and Dempster-Shafer calculi appear to offer the most mathematically sound procedures for manipulating and combining numerical degrees of belief in propositions that are either true or false. Because of this, these methods were selected from the many in existence for the development of classifiers for biological classification.

Chapter 4
Expert Knowledge Elicitation

4.1 Introduction

This chapter describes the work undertaken to elicit domain knowledge from a leading expert in the biological assessment of freshwaters. Interviews with the expert are described in which basic domain concepts are elicited, followed by the selection of key indicator taxa and the identification of abundance-levels as sensor states. A novel technique for eliciting probabilistic knowledge corresponding to a taxon's indication of river water quality is described in detail. The theory and practice of knowledge acquisition, and particularly the elicitation of measures of uncertainty, are reviewed.

4.2 Overview of Knowledge Acquisition and Elicitation

Knowledge engineering, the essential process of building an expert system, involves the acquisition of knowledge, its representation and the selection of inference procedures for manipulating that knowledge within a computer system. The knowledge encoded in expert systems may be the expertise from a human expert, or incorporate general knowledge from the literature. Alternatively, inducing rules implicit in data may be possible, a process known as machine learning or induction.

4.2.1 Machine Induction

Knowledge engineering is an expensive and time-consuming activity, a fact that provides the motivating force behind research into machine learning. During the 1980s, research in artificial intelligence focused on the possibility that computers could somehow synthesise knowledge themselves. In machine learning, the knowledge base for the domain is derived from the analysis of a large set of examples that essentially train the system.

A prominent example of a working system based on induced rules was that built by Michalski and Chilausky (1980) for identifying diseases in soybeans. The inductive algorithm they developed, AQ11, was presented with a training set of examples consisting of detailed descriptions of the soybean plants and the expert's diagnosis of their condition. The resulting expert system was extremely successful in correctly diagnosing new and unseen samples, and in fact surpassed the expert in performance.

The ID3[1] algorithm also uses a training set of examples, described by a series of attributes and resulting classes, to induce IF...THEN rules for a decision tree. Examples that are not in the training set can then be classified using the derived rules. Selection of the attributes and examples is crucial to the procedure: a poorly-derived training set may lead to invalid rules. An attribute is selected by the algorithm to subdivide the training set, creating an intermediate node in the decision tree corresponding to a rule. The algorithm continues iteratively until the tree is complete. A detailed description of ID3 is given by Michie (1979); a good example is presented by Hart (1989).

Neural networks belong to this class of machine learning methods, but differ in that knowledge is stored as connection weights between nodes, rather than being represented by any explicit form such as production rules or frames. In these so-called connectionist machines, expertise is distributed throughout the network rather than stored locally (McLelland and Rummelhart, 1986). Neural nets adjust their internal connections in response to example inputs and required output. When properly trained, the nets can output correct responses to new input signals. The potential of neural nets lies in their ability to process massive amounts of sensory input data in applications for which algorithmic solutions do not exist, and their ability to model relationships between data that are complex enough to preclude a rule-based approach (Giarratano and Riley, 1989).

4.2.2 Knowledge elicitation

Although there are limits to what the knowledge engineer can learn from written materials (Open University, 1989), background reading can provide a basic vocabulary of domain concepts. This literature-based acquisition from documents, such as books, patents, articles, and reports is often a prerequisite to discussions between the knowledge engineer and the expert. It is however the process of acquiring knowledge directly from the expert, usually known as knowledge elicitation, which remains the primary method.

Some techniques available for eliciting knowledge from an expert include the direct methods of interviews and protocol analysis, and the indirect methods, in which knowledge is inferred, of the laddered grid, card sort, and repertory grid. Interviews are intended to reveal the expert's ideas, the relationships between concepts and their organisation, and the reasoning processes employed for judgement and problem-solving (Open University, 1989).

[1] Iterative Dichotomiser 3.

According to Cullen and Bryman (1988) the most common technique is that of the unstructured interview. These are useful for defining the problem or deciding its scope, providing suggestions for questions at later sessions. Recording interviews on audiotape is usual; in some situations the sessions are filmed. The interviews are then transcribed for later analysis. Such interviews have the advantage that the domain expert is more likely to feel at ease in the relaxed environment furnished by the informal or unstructured interview: the disadvantages include the production of extensive transcripts and the difficulty, from the knowledge engineer's viewpoint, of controlling the conversation.

Structured or focused interviews allow rules and ideas to be extracted more efficiently than the conversational type of interview. Focused discussions are normally conducted by following some agreed agenda (Ahmad and Griffin, 1991). They can be viewed as setting a task for the expert, requiring him or her to produce a verbal report, supported perhaps by diagrams, sketches and so forth.

The 'think aloud' protocol is a means by which the strategy adopted by people solving actual problems can be studied. While solving some familiar problem, the subject gives a verbal presentation that is recorded, and analysed in detail. The resulting *protocol analysis* for well-defined problems derives a model of the problem about the initial state, the goal state, and legal operations through the 'state space' to achieve the goal. Analysing protocols is extremely difficult, even for very simple problems. For ill-defined problems (such as those that occur in medical diagnosis), the difficulties are compounded (Open University, 1989).

Indirect methods derived from psychological studies include: - '*twenty questions*', in which the knowledge engineer suggests some situation and tells the expert to ask questions about this case, *laddered grids* used to elicit data for hierarchically-structured domains such as taxonomies, and *card sort*, in which cards bearing the names of domain objects are sorted according to whatever criteria the expert considers appropriate (Hart, 1989). These methods aim to 'bypass the cognitive defences'; they can however, seem patronising to the domain expert (AIAI, 1990).

Burton *et al.* (1990) have evaluated four knowledge elicitation techniques for classification domains: the structured interview, protocol analysis, card sort and laddered grid. Protocol analysis was the least effective technique. The two indirect techniques, card sort and laddered grids were useful in providing complementary knowledge to the standard interview.

4.3 Theory and Methods of Uncertainty Elicitation

4.3.1 Studies in uncertainty estimation

Several authors have examined both the psychological and practical aspects of eliciting measures of uncertainty. Many studies have been made of calibration: how close an elicited probability is to its objective value. Weather forecasters for instance, are highly calibrated in their estimates (Winkler and Murphy, 1968). According to Tonn and Goeltz (1992) a good probability estimator is one who is both highly calibrated and reliable, i.e. gives consistent uncertainty measures over a time for an identical set of questions. Note that a reliable probability assessor can be miscalibrated, i.e. measures may consistently be under- or overestimated.

To arrive at these estimates, it is believed that people engage in some form of cognitive problem-solving behaviour. When asked to assess a probability of some event, the assessor may apply some mental model of the domain knowledge or use a so-called intuitive approach, in which a measure is given with little conscious thought (Tonn and Goeltz, 1992). The elicited measure may be in a form that requires transformation or adjustment to one that can be used in the application intended by the knowledge engineer. For instance, the assessor may show the probability measure by some mark on a line rather than a number, according to the elicitation exercise. This may then have to be transformed by some means specific to the exercise into (say) a probability or odds-value. One graphical approach is the "probability wheel" in which a probability is obtained "without explicitly mentioning a number" (Henrion *et al.*, 1991). An adjustment may however still be required even if the elicited measure is in numerical form.

Another way in which people derive probability measures is to use the so-called analytic approach involving the use of heuristics in problem-solving. This can take place in several ways. The assessor may draw analogies between the posed problem and another (solved) problem, or perhaps has some quantitative knowledge of the event (such as its frequency of occurrence). Thirdly, the assessor may imagine situations by which the event under investigation becomes true. An interesting psychological effect of this process is that the more easily an event can be imagined, the higher its assessed likelihood will be (Tversky and Kahneman, 1974). This effect was observed during the elicitation exercise carried out for this project. For instance, the expert often ruled out events as impossible that could perhaps be more correctly termed has highly improbable. The use of zero or very small probability measures has implications for uncertain reasoning calculi (see for instance

Dubois and Prade (1985)).

These and other observations motivated Tonn and Goeltz (1992) to investigate several hypotheses regarding probability assessment and reliability. These include the conjecture that the method or the 'measurement scale'[2] which a person uses to report an uncertainty estimate will affect its reliability, since a method that promotes an analytic approach to assessment will be inherently more reliable than one in which a measure is given intuitively. Secondly, changing the mode of assessment (e.g. from giving 'chances' or 'odds' to making a mark on some scale) may adversely affect the reliability, since different problem-solving approaches may be employed. This suggests that a knowledge engineer should follow a consistent elicitation method or use those that involve the same problem-solving approach for the expert.

Tonn and Goeltz carried out a comprehensive investigation of these and other hypotheses by asking questions relating to the occurrence of a wide range of events from a large sample size of subjects. The subjects were allowed to answer the questions against three measurement scales: probabilities, chances, and percent of the time. The findings support the conjectures regarding the modes of uncertainty estimates, and the hypothesis that estimates for highly probable events are more reliable. Lastly, they suggest that the same elicitation methods be used with the same set of questions. Tonn and Goeltz's findings lend support to the methodology employed for the practical elicitation of uncertainty estimates for the biological classification problem, discussed below.

4.3.2 Words versus numbers

One aspect of the uncertainty assessment problem not directly investigated by Tonn and Goeltz is the question of whether words or numbers should be used in communication. Bryant and Norman (1980) and Nakao and Axelrod (1988) in two separate studies investigated the use of probability estimates in medical science. The studies were motivated by concerns that serious misunderstanding between professionals may occur in the communication of words such as 'probable', 'sometimes', 'likely', 'excludes', 'typical', and so on, which are intended to convey estimates of the probability of some condition. Bryant and Norman asked physicians to associate numbers (from zero to one) with a list of 30 expressions of probability. Expressions of low probability had very large ranges due to the

---

[2] Tonn and Goeltz refer to this as the "answer modality".

wide variation in the numbers assigned, while only nine expressions, each associated with high probability of disease, had ranges lower than 0.5. From this study the authors assert that only numerical probabilities should be used in communication between medical professionals.

Nakao and Axelrod carried out a similar study using a large sample size of respondents, and from their findings insist that "verbal expressions of frequency should be eliminated from medical communications". Budescu *et al.* (1988) in an analysis of decisions based on numerical and verbally expressed uncertainties similarly concluded that "numerical judgements were significantly superior".

4.3.3 Implications for elicitation of probability distributions

This section presents a brief summary of the above review and its implications for the work carried out in this project. Results from Tonn and Goeltz's investigations suggest that a consistent method for eliciting measures of uncertainty should be adopted; otherwise, the reliability of the measures will be affected. A method that promotes an analytical approach to uncertainty estimation will be more reliable than an intuitive approach. Several authors caution against the use of "imprecise" measures of probability, recommending the use of numbers rather than words. As discussed in Chapter 3, non-numeric methods of uncertain reasoning are considered inadequate to handle imprecise information: for this, confidence levels or degrees of support are required. Therefore, uncertainty elicitation for numerical reasoning schemes must eventually produce numerical values.

However, probability measures may be elicited via a non-numeric approach, which would then need to be transformed to numbers if necessary. The knowledge engineer should be aware however of psychological bias exerted by uncertainty estimators. Events that can be easily imagined often receive higher assessed likelihoods, whereas events that might more correctly be described as highly improbable are estimated as impossible.

Later sections in this Chapter describe the development of a graphical method for eliciting uncertain benthic knowledge as conditional probabilities. The scheme adopts many recommended approaches discussed here.

4.4 Interviews with the Expert

4.4.1 Interview arrangements

The knowledge elicitation for this project was conducted as a series of seven interviews between the domain expert (H.A. Hawkes) and in the role of knowledge engineers W.J. Walley and the author. The interviews themselves evolved from the more conversational, unstructured type in the early sessions to those with a more definite focus as particular ideas were explored. The intention of the early interviews was to gain an overview of the domain ideas, before turning to the problem of eliciting numerical estimates of probability in later interviews.

All the interview sessions were recorded onto audiotape to allow transcription and analysis by the author, and to ensure that information revealed during the sessions was not later overlooked by the interviewers. Each interview absorbed most of a working day, being divided into a morning and afternoon session. This arrangement, while clearly demanding of the expert's time, meant that the interviews themselves were relaxed and non-intensive.

In transcribing taped interviews to word-processed text, a decision had to be made regarding how much detail to include in the transcriptions. For the more conversational discussions, a great deal of transcript is created if the dialogue is reproduced word for word. Only a small percentage of the transcripts will contain core domain ideas; most of the conversation involves setting the context for a particular question to be put to the expert. However, transcribing only part of the conversations to text runs the risk of omitting important knowledge contained in asides or issues raised apparently out of context, which with hindsight should have been included. The transcriptions were therefore largely verbatim, so that the process took in the order of two to three days or more to complete for the average interview. Transcript length for the more detailed interviews was approximately twenty or more pages.

Where possible copies of the transcript of the previous interview were presented to the participants at an interview session, to provide a starting point for discussion. Alternatively, reports relating to biological surveillance, or lists of benthic taxa would be used as a catalyst for discussion, in cases where an agenda for the meeting had not been decided beforehand. In the fifth interview for example, an interim report written by the author on his understanding of domain concepts was the subject of a detailed response by the expert.

Besides this verbal elicitation of knowledge, the expert contributed notes, reports, drawings of benthic invertebrates, annotated examples of biological classification, and most notably, graphical depictions of the probability distributions for the indicator taxa. The elicitation of this graphical data is described in detail below. The author also took part in a one-day introductory practical session on biological surveillance in the field organised by Mr. Hawkes.

In the following section extracts from the interview sessions are used to illustrate how domain ideas central to this project were elicited from the expert. Given the length and number of the transcripts, this is necessarily a brief overview. Initials WJW and MB refer to comments or questions from the knowledge engineers W.J. Walley and the author while HAH refers to the domain expert Mr. H.A. Hawkes. Where not stated, the quotation is from the expert.

4.4.2 Elicitation of domain concepts

4.4.2.1 Meaning of "water quality"

The participants returned to this issue over several interviews, since it was problematic. In one session the expert outlined his view of the distinction between "river quality" and "water quality" in his work with the Biological Monitoring Working Party to produce the BMWP score system:

"Our terms of reference were first set up by the DoE were [...] in fact to produce a biological method of surveillance which would provide data on water quality, and, as a group we've rejected this ... what we would do [was to] produce a system of biological monitoring that would monitor river quality, which is something different from water quality. River quality is more than water quality ... because you've got water flowing down a river, engineers come along and put it in a concrete bottom, and all sorts of other things like that: it alters the quality of the river, markedly, without altering the water quality at all ... so there are subtle differences between river quality and water quality."

Or again, in the context of organic pollution:

MB: "Do saprobic zones refer to water quality or river quality?"
HAH: "Water quality generally, but an organic discharge can affect the substratum and thereby affect river quality."

The expert also dealt with the relation between sampling effort and the nature of the survey in question:

"I distinguish between what I call environmental pollution - that's the effect it has on the environment generally, the aquatic environment, in its effect on the organisms there, and the other pollution which affects water quality in relation to man's use of that water ... [For instance] you may find that some stoneflies have disappeared from the river but you can still use it for water supply! The two are not incompatible - but it's a matter of degree. For the environmental impact, you really need a much more specific and detailed study than you do for just getting a BMWP score, which tells you what you need to do about your water quality."

However, the differentiation between "river water quality" and "water quality" continued to perplex this author throughout the project, and prompted an investigation of the literature for clarification. The issues were discussed in some detail in Chapter 2.

4.4.2.2 Discrete classification and saprobic valency

The evolution of discrete classification systems for biological water quality has been described in Chapter 2 of this thesis. Within the knowledge acquisition sessions, frequent reference was made to the correspondence of the saprobic zones to gradations of organic pollution, and to the existence of discrete biological classification systems. Reports of the biological quality of rivers in the Yorkshire area, dating from the late 1970s, were one of the subjects of discussion at the second interview. The reports contained examples of the direct biological classification of benthic samples, in terms of classes B1a,B1b,B2,B3,B4. It was apparent during discussions that there was also a correspondence between saprobic zones and discrete biological classes. The idea of saprobic valency, which shows the likelihood of finding a species in each of the five saprobic zones, is seen to parallel closely that of numerical belief in the biological classes.

The expert considered the saprobic system useful, but was sometimes concerned about the assignment of certain values to particular species, and its applicability to British waters. Referring to the example of the mayfly *Baetis rhodani*, the expert disagrees with the saprobic values assigned to it by Continental workers:

"*Baetis rhodani* [which] is an exceptional mayfly - more tolerant of organic pollution than anything else ... on the saprobic classification is shown to indicate the best-quality waters - xeno- and oligo-saprobic."

Despite these reservations, the expert undoubtedly appealed to the system when deriving the probability distributions for the indicator taxa.[16] Observation of this connection during the elicitation exercise led to a suggestion by the author that the saprobic valencies should be used directly to obtain the conditional probabilities $P(H|e)$. This suggestion was not pursued for several reasons. Although the saprobic system is one that has considerable support in continental Europe it is less popular in Britain. It is based entirely on species-level data, with no information for taxa at family level or higher. Since identification to species level is often onerous or impossible, it was felt that it would be undesirable to rely on an exclusively species-based system. This could preclude the use of data on invertebrates

[16] The domain expert had previously used the saprobic system in the biological surveillance of Midlands rivers (Hawkes, 1956).

at higher taxonomic levels should the indicator group be enlarged in the future. Moreover the valencies were designed to be used with a system whose method of combining evidence is essentially *ad hoc*, in contrast to numerical methods of uncertain reasoning that have sound mathematical foundations. A prime consideration in the project was to acquire knowledge from an acknowledged expert in the domain as directly as possible, rather than appeal to the literature.

4.4.2.3 Importance of abundance levels

The significance of invertebrate abundance in benthic samples was a point emphasised by the expert on several occasions. For instance, referring to the numbers of a particular taxon in a sample:

"My appreciation of a sample certainly involves a rough abundance. If there's only one there, it tells you something quite different than if there's a lot there."

The expert distinguished levels of abundance:

"It depends on how you're going to use this abundance. There are two grades: significantly present as I call it, and abundant."

This distinction partly justified the refinement of taxa states into "Established" and "Abundant" (discussed below). Also, the term 'abundant' applied to a sample implies different numbers of individuals for different taxa:

" ... with abundance you got to have different values for different taxa ... for Tubificid 'abundant' means in terms of thousands ... where for Planariidae it means units ... if I've got 10 in my sample they're quite abundant ... in relation to what they are in anywhere else ... if I get Tubificid worms in several 100 that's ... quite common ... so abundance is not just a numerical value"

Within this context the question of evidence suggested by taxa that are absent from a sample was also discussed:

"I do deal with the absence of things, I look at the things that are present, and what they tell me, and then I start looking for absences ..."

However, 'absent evidence' is not as important as that provided from those taxa present, and evidence may have different weighting:

"One thing I would emphasise is that the appearance of things that should be there outweighs the absence ... if things are absent, then you may not have caught them, but if they're there, and they shouldn't be there, then these are very strong indicators to me. *Asellus* and leeches in a riffle are very strong indicators of organic pollution, more than the absence of some species of mayfly."

4.4.2.4 Selection of Benthic Indicator groups

During an interview session a comprehensive list of all the major benthic invertebrates previously drawn up was closely examined to identify 'key' indicator taxa: those invertebrates considered by the domain expert to be of particular value for determining biological water quality. This reduction was necessary for several reasons. It was agreed early in the project that the task of eliciting detailed ecological knowledge on all these invertebrates was impractical given the size of the database. Knowledge on the more rare or exotic invertebrates was either scarce or absent, and in any case might be of limited worth for assessing biological quality from sample inspection, since the biota in question may (for example) have low indicator value. So it was decided to focus on those that were generally of common occurrence and were of high indicator value.

The selection of small subsets of the invertebrates aided understanding more fully how the domain expert assessed biological quality from his knowledge of the ecological requirements of the invertebrates. Various subsets of the complete list were drawn up in collaboration with the domain expert who then provided his own detailed descriptive knowledge of the benthic ecology for each taxon. This was to prove a valuable precursor for the uncertainty estimates later elicited for the classification problem. The following extract shows some of the detail obtained:

WJW: "We've talked about getting 10 species that you can talk about - about which you can tell us something very specific... we'd like to know all there is to know about them."

HAH: "I wrote down as much as I could about each of them .. is this the sort of information you want? [Shows handwritten notes prepared]. This is in roughly the order of sensitivity to pollution... *Leuctra fusca* is a stonefly, and therefore it's very sensitive - all stoneflies are, that's why I think it's profitable to look at stoneflies. If I've got stoneflies present, in my stream, I don't need to look much further because it tells me straight away it's pretty good, and if the other species there don't tell me that, then there's some explaining to do, but normally I'd expect *Gammarus* to be present, and caddis, along with *Leuctra*.

The presence of *Leuctra fusca* specifically, being a stonefly, is indicative of good quality water, but *Leuctra fusca* is probably one of the most pollution-tolerant, so it's at the 'edge of the range' - still very sensitive to pollution, but slightly more tolerant.

Plecoptera are the most sensitive insect order, to organic pollution, usually associated with upland eroding substrata streams, such streams as stenothermal (low temperature range), oligotrophic (nutrient-poor, well aerated), that's where you'll find these ... but out of the stoneflies this species *Leuctra fusca* is probably one of the most tolerant to organic pollution. But then it's given a saprobic value of 2.15 - I don't believe that - that's more tolerant than *Gammarus*. It's too high a value in my experience in the U.K. Interestingly, according to the FBA, *Leuctra fusca* was found as far downstream into lowland rivers as any stonefly, and looking at the distribution of species in this list, the more tolerant they are of organic pollution, the more likely they are to extend into a lowland river, because they you'll get low oxygen."

These descriptions related to the first indicator group selected, which comprised ten species, selected by the domain expert on the basis on their indicator value and their "spread" across the range of water qualities from good to poor. The expert volunteered specific details on taxa within the group, for example *Simulium ornatum*:

HAH: "... unlike the others which are dominated by water quality it depends on the bacterial numbers present in the water, because it sits in the water and strains off the plankton, so water that it rich in plankton encourages *Simulium ornatum*, and you find in outflows from eutrophic lakes. The first thing I'd say about it is it wouldn't worry me if I didn't find any - because its occurrence isn't that frequent, so its absence isn't all that significant. On the other hand, in some situations you can get your net full of them, it's in the B2 or B3 range.

MB: "Why did you include this particular taxon in this small group?"

HAH: "It tells me something about the planktonic concentration of the water, not necessarily in relation to water quality, although it is classified on the saprobic system"

The domain expert supplemented the verbal descriptions with handwritten notes on each of the taxa, which are given in the next section.

4.4.2.5 Elicited knowledge on first ten indicator taxa

The following edited descriptions are in note form following the originals supplied by the domain expert:

(1) *Leuctra fusca*: Plecoptera - Leuctridae (BMWP score 10)
Plecoptera is the insect order most sensitive to organic pollution. Usually associated with upland eroding substratum streams, stenothermal, oligotrophic (nutrient-poor), well aerated. *L. fusca* is probably one of the most tolerant of the stoneflies to organic pollution. (Quoted saprobic index of 2.15 is probably too high a value for British conditions). Interesting that according to FBA *L. fusca* was found as far downstream into lowland rivers as any stonefly. [Low oxygen in lowland rivers]. Presence in lowland streams is an indication of good conditions, absence of organic pollution. River Ystwyth - *Leuctra* found to be <u>more tolerant</u> of metal mine waters (Pb & Zn) than other taxa: crustacea, oligochaeta, leeches & molluscs.

(2) *Rhyacophila dorsalis*: Trichoptera (Caddis) - Rhyacophilidae (BMWP score 7).
A non-cased caddis. Rhyacophilidae - restriction to fast-flowing waters 80-90 cm/sec. - especially in moss on rocks. On stones without moss, restricted to lower surfaces. Hence, most common in upland streams. Rhyacophilidae is sensitive to organic pollution - *R. dorsalis* probably most tolerant of the family and found in mildly organically enriched waters but not in waters with any high degree of pollution. Tolerant of high ammonia concentration. Distinct from net-spinning caddis (e.g. *Hydropsyche*), therefore less affected by suspended solids. Active foragers - predators.

(3) *Hydropsyche angustipennis*: Hydropsychidae (BMWP score 5).
A net-spinning caddis, and therefore susceptible to suspended solids in water, e.g. coal-dust.
Family Hydropsychidae - found in flowing waters - (net feeding) - succession of species of *Hydropsyche* downstream but *Hydropsyche angustipennis* not in this sequence. Sometimes found in lower reaches of large rivers but also in small streams below outfalls from ponds and lakes. Known tolerance to high temperatures, low oxygen & low water velocities. Known to increase in numbers in a fair degree of organic enrichment to assume large populations - most tolerant by far of Hydropsychidae. This species is probably the reason for the low BMWP score for family. Tolerant of high ammonia (as other caddis).

(4) *Gammarus pulex*: Gammaridae (BMWP score 6) - Crustacea - Arthropoda.
Freshwater Shrimp. One of the most commonly occurring invertebrates in streams. Typical in riffles and displaced by the other common crustacean *Asellus aquaticus* with increasing organic pollution. Sensitive to quite low oxygen levels, but oxygen pattern determines presence or absence. *Gammarus pulex* is very sensitive to ammonia. Thus below polluting sewage effluents low oxygen and high ammonia concentration bring about a reduction and eventual elimination of the population. Under such conditions *Asellus* increases in riffles; hence the proposed *Gammarus/Asellus* ratio as a measure of organic pollution. Saprobic Index - 0.65. Tolerant of some toxicants (insecticides) at concentrations suppressing insect populations. Found in waters of moderately high to low mineral content but never where there is any brackish water influence, e.g. estuaries or mineral springs - Not in waters consistently below 5.7 pH.

(5) *Simulium ornatum*: Simulidae (BMWP score 5) - Diptera Larva.
Blackflies. Most tolerant of insects to strong currents - requires current for feeding by straining food from water flowing over them > 1.2 m/sec (*S. damnosum*). Encouraged by mildly polluted conditions because of the increased food available in the nutrient-rich water but where pollution results in heavy slime growths on stones they are reduced in numbers and eventually eliminated. Needs surfaces for attachment - rocks or vegetation. When conditions favour the species, assumes large populations. Adult *Simulium* flies are a nuisance for man and cattle by biting. In Africa, they are a vector of Onchocercus - African River Blindness.

(6) *Lymnaea peregra*: Lymnaeidae (BMWP score 3) - Mollusca
Wandering snail. Ubiquitous - probably the most widely distributed invertebrate - from upland to lowland rivers - more common in lowland rivers. Although found in good quality streams, also found in moderate-to-high degrees of organic pollution. Saprobic Index 2.0 ($I = 2$). Most abundant in hard waters but also present in soft waters. Suited by moderate currents rather than very strong ones. Also found in lentic waters - lakes.

(7) *Asellus aquaticus*: Asellidae (BMWP score 3) - Crustacea
Water Hog-louse. Tolerant of moderately low oxygen and high ammonia concentrations. Rare or absent from riffles of good quality streams but may be found in pools in bottom muds. A scavenger, therefore appearance in riffles a strong indication of organic pollution. Being a scavenger feeding at a low trophic level may assume large populations under suitable conditions of organic enrichment. More common in lowland rivers.

(8) *Erpobdella octoculata*: Erpobdellidae (BMWP score 3) - Hirudinea
Leeches. Saprobic Index 3.0. Occurs in a wide range of habitats - lotic & lentic, hard water & soft water. Most abundant in moderately organically polluted situations. Widely distributed in Britain. Feeds on insect larvae (chironomids) & worms. Rare or absent in good quality trout streams, probably because of scarcity of food organisms or because of predation by fish. Invades riffles with increasing organic pollution (Saprophilic), only eliminated in very organically polluted conditions (e.g. Polysaprobic). More common in lowland rivers. Require solid hard surfaces for attachment.

(9) *Chironomus riparius*: Chironomidae (BMWP score 2) - Diptera
Red blood-worms. By virtue of possessing haemoglobin, the larvae are able to exist in very low oxygen concentrations; experimental evidence suggests they benefit from lower oxygen levels. It feeds by straining micro-organisms including bacteria from the water. It therefore thrives in organically polluted streams with low oxygen and high bacterial numbers. It is found most abundantly in organically polluted riffles on silted streams and stream beds. It invades riffles which are organically polluted, not being found in non-polluted riffles. Saprobic value 3.65. Its presence in riffles is a strong indication of severe organic pollution. It is however affected by insecticidal toxicants. It has been used in bioassay studies on insecticides in sewage effluents after treating the filters with insecticides to control sewage filter flies.

(10) *Tubifex tubifex*: Tubificidae (BMWP score 1).
Sludge worms. Live in bottom muds of rivers in which they construct tubes. Possess haemoglobin and like *C. riparius* able to live in low oxygen conditions. They feed on the organic matter in the mud by ingestion. They are therefore found most abundantly in organically polluted streams with low oxygen and where the bed has been covered with an organically-rich deposit. Although most abundant in organically-polluted streams they may be found in less polluted streams in smaller numbers. In natural (non-polluted) streams they are most commonly found in the lowland rivers - presumably because of the depositing substratum. Saprobic index 3.8.

4.4.3 Elicitation of probability distributions

4.4.3.1 Overview

Since biological classification involves uncertain reasoning, numerical estimates of the probabilities associated with the decision problem were sought from the expert. The fourth and particularly the fifth interviews were concerned with this elicitation. This section describes the results of this work, and where appropriate illustrates the elicitation process via extracts from the discussions. The uncertainty measures which were initially sought from the expert refer to the occurrence of the first ten indicator taxa in *riffles*, generally fast-flowing regions of the river with stony substrata.

4.4.3.2 Initial elicitation

Given the dominance of Bayesian analysis in uncertain reasoning, the simplified calculus was considered first. Whatever form of the simplified method is used (odds-likelihood or multihypothesis), the conditional probabilities must be directly or indirectly elicited in some form from the expert. If the odds-likelihood form of Bayes' rule is used, the coefficients of logical sufficiency and necessity L_s and L_n can be sought directly. However, it was decided not to pursue this approach for several reasons. In spite of arguments that the language of odds gives meaning to probabilistic ideas, it was found from discussions with the expert that he was no more comfortable with odds than probabilities. Moreover the distinctions between "chance", "likelihood" and "odds" and similar terms were not clear. Although an elicitation exercise could have been devised which hid these distinctions from the expert, there was some apprehension that the uncertainty estimate so elicited could have had an ambiguous interpretation.

Given these arguments and the fact that odds translate to probabilities directly (and vice versa) it was determined to maintain a consistent terminology of "likelihood", "chance" or "probability" as referring to probability values in the elicitation exercises. The question arises which of the two forms of conditional probabilities should be supplied. According to Shacter and Heckerman (1987) since $P(e|H)$ represents causal knowledge it is easier to elicit in diagnostic systems than $P(H|e)$. This is particularly true if there are many hypotheses H_i. Thus eliciting the probability of observing the symptom given that the disease is present is easier rather than the converse: if the symptom is observed (e), it may be caused by many diseases H_i, whereas a particular disease may produce a specific symptom. In Bayesian updating the posterior probabilities are obtained from the inverse formulation as $P(e|H)$, and these values are normally sought from the expert (Pearl, 1987).

However Dillard (1992) points out that some evidence contributes more naturally to $P(H|e)$. In discussions with the domain expert it was found that this was true in this domain: possibly because there are only a small number of competing hypotheses ($n = 5$). This was discovered only after some experimentation. Initially both $P(H|e)$ and $P(e|H)$ were elicited for the ten indicator taxa to compare values and to find out in which mode the expert preferred to express conditional probabilities.

It was felt that requiring the expert to express numerical values directly was unreasonable. The expert was therefore asked to associate the verbal expressions "very likely", "likely", "possibly", "unlikely", "highly unlikely" across the range of five water quality classes for the occurrence therein of each of the ten taxa. These expressions were later transformed into numerical values in a manner explained below. To simplify the initial elicitation, the benthic states were considered to exist in two states only: presence and absence (i.e. e and $\neg e$). Probability values for absence were derived from values supplied for the present state, for which the expert's judgement was sought for the two modes.

To elicit the two numerical judgements of $\{H_i\}$ the expert was asked questions in the following two forms:

(1) given that the evidence (i.e. a particular taxon) is present at a river site, how likely is it that the river water-quality class is (say) B1a?

and

(2) given that the water quality is of class (say) B1a?, what is the likelihood that the taxon is present/well-established?

For each format, the question is asked for each of the five quality classes. Questions in the form of (1) clearly elicit $P(H|e)$ while (2) elicits $P(e|H)$. The elicitation of these two modes for the ten taxa resulted in the verbal expressions given in **Tables 4.1** and **4.2**.

Table 4.1 Initial elicitation of $P(H|e)$ using verbal expressions

Taxon	**B1a**	**B1b**	**B2**	**B3**	**B4**
Leuctra fusca	L	P	U	HU	HU
Rhyacophila dorsalis	P	L	P	HU	HU
Hydropsyche angustipennis	HU	U	L	U	HU
Gammarus pulex	L	VL	P	HU	HU
Simulium ornatum	HU	U	P	P	HU
Lymnaea peregra	U	P	L	L	U
Asellus aquaticus	HU	HU	L	L	U
Erpobdella octoculata	HU	HU	L	L	P
Chironomus riparius	HU	HU	U	P	L
Tubifex tubifex	HU	HU	U	L	VL

Key: HU - highly unlikely; U - unlikely; P - possibly; L - likely; VL - very likely

The likelihoods were scrutinised for their correspondence with saprobic valencies in the fifth discussion. Although a few indicators showed a close correspondence with their equivalent valencies, the differences were significant enough to suggest that the distributions were independent of the equivalent saprobic valencies. Valencies were not in fact available for two of the indicators: *Rhyacophila dorsalis* and *Hydropsyche angustipennis*.

These verbal expressions were converted to numerical values by arbitrarily assigning an interval to each within the range 0 - 100 as follows: highly unlikely - 0:5, unlikely - 5:25, possibly - 25:55, likely - 55:75, very likely - 75:100. Median values within each interval were then assigned to the $\{H_i\}$, i.e. $P(\text{"HU"}) = 0.025$, $P(\text{"U"}) = 0.15$, $P(\text{"P"}) = 0.4$, $P(\text{"L"}) = 0.65$, $P(\text{"VL"}) = 0.875$. Using this scheme the verbally-expressed probabilities $P(e|H)$ for *Leuctra fusca* become $\{0.875, 0.65, 0.15, 0.025, 0.025\}$. Values for $P(H|e)$ were normalised to sum to unity.

Table 4.2 Initial elicitation of $P(e|H)$ using verbal expressions

Taxon	**B1a**	**B1b**	**B2**	**B3**	**B4**
Leuctra fusca	VL	L	U	HU	HU
Rhyacophila dorsalis	L	VL	P	HU	HU
Hydropsyche angustipennis	U	U	L	P	HU
Gammarus pulex	L	VL	P	HU	HU
Simulium ornatum	HU	U	L	P	HU
Lymnaea peregra	U	P	L	L	U
Asellus aquaticus	HU	HU	L	L	U
Erpobdella octoculata	HU	HU	L	VL	L
Chironomus riparius	HU	HU	L	VL	L
Tubifex tubifex	HU	HU	U	L	VL

Key: HU - highly unlikely; U - unlikely; P - possibly; L - likely; VL - very likely

For the odds-likelihood formulation of Bayes' rule, used in the early classification systems developed in this project, the values of L_s and L_n must be calculated from the supplied conditional probabilities. Ideally, the unknown probability $P(e|\neg H)$ could be supplied by the expert. However, the form of the required question: "what is the chance of finding the evidence given that the quality class is *not* (say) {B1b}?" was considered too obscure for the expert to supply a meaningful uncertainty estimate. This probability value can in any case be determined from the $P(e|H_i)$ for each piece of evidence from the relation

$$P(e|\neg H_i) = \frac{\sum_{j \neq i} P(e|H_j)\,P(H_j)}{P(\neg H_i)} \tag{4.1}$$

from which can be derived L_s and L_n since

$$L_n = \frac{P(\neg e|H)}{P(\neg e|\neg H)} = \frac{1 - P(e|H)}{1 - P(e|\neg H)} \tag{4.2}$$

Using this approach of course avoids the problem in the simplified formulation which can arise from inconsistent assignments of L_s and L_n. Applying these equations to the numerical values for the probabilities resulted in L_s and L_n figures as shown in **Table 4.3**. The fact that most of the L_n values are non-unity suggests that the absence of taxa is significant.

Table 4.3 L_s and L_n for first ten taxa derived from verbal expressions

Taxon		B1a	B1b	B2	B3	B4
Leuctra fusca	L_s	4.12	2.418	0.38	0.058	0.058
	L_n	0.16	0.47	1.402	1.69	1.69
Rhyacophila dorsalis	L_s	1.96	3.18	1.01	0.09	0.09
	L_n	0.52	0.17	0.98	1.90	1.90
Hydropsyche angustipennis	L_s	0.48	0.48	3.58	1.64	0.074
	L_n	1.225	1.225	0.42	0.79	1.47
Gammarus pulex	L_s	1.96	3.18	1.01	0.09	0.09
	L_n	0.52	0.172	0.98	1.90	1.90
Simulium ornatum	L_s	0.08	0.545	4.33	1.88	0.08
	L_n	1.40	1.17	0.41	0.76	1.40
Lymnaea peregra	L_s	0.32	1	1.92	1.92	1.92
	L_n	1.58	1	0.52	0.52	1.58
Asellus aquaticus	L_s	0.067	0.067	3.05	3.05	0.44
	L_n	1.54	1.54	0.44	0.44	1.28
Erpobdella octoculata	L_s	0.045	0.045	1.65	2.59	1.65
	L_n	2.16	2.16	0.57	0.18	0.57
Chironomus riparius	L_s	0.054	0.35	0.35	2.16	3.58
	L_n	1.79	1.47	1.47	0.5	0.21
Tubifex tubifex	L_s	0.048	0.30	0.94	1.79	2.85
	L_n	2.02	1.65	1.04	0.54	0.18

4.4.3.3 Evaluation of Initial Elicitation

The efficacy of the initial elicitation was reviewed during a particular interview session. The expert remarked on the difficulties of precise definitions of biological classes, and his use of biotic indices to confirm his own assessment of quality:

"I've got to remind myself what a B3 and a B4 means ... it's outside my experience ...most of my surveillance work ... when I get my survey data, I work out the indexes as a guide but I look at the data and I immediately compare it with the data I had before and I'm fairly happy to say whether a thing's got better or worse. How much better or worse, then I sometimes have to refer to the indices. I could show you the Don data - last time there was only one species of mayfly at this station, but this time there are three species of mayfly - significantly present - this represents an improvement since last year. This is confirmed by the increasing values of the TBI, biotic scores ..."

Also, his appeal to the saprobic system did not always help the process of biological classification. The biological classes imply an increasing degree of organic enrichment from B1a to B4, but a river biologist normally observes the response of the benthic communities as levels of organic material in the river decrease due to degradation. The following extract illustrates his difficulty when attempting to state the "preferred class" of *Asellus aquaticus*:

"Let me think ... take out B1a and B1b, certainly not B4 ... the reason I'm hesitating [is that] I've got to reverse things in my mind all the time - I'm used to taking organic pollution in the stages of recovery ... you've got it the other way round!"

Despite these reservations, the expert became comfortable with the use of the discrete classes after classifying several sets sample data, as described in Chapter 5.

The verbal expressions of likelihood were discussed with the expert to ascertain that he agreed with the probabilities derived from them. Some modification took place on this basis. For instance for *Lymnaea peregra*:

MB: "For *Lymnaea peregra* the valency is 0 3 4 3 0 - the distribution is "U,P,L,LU "and we calculated that as 0.1,0.2,0.3,0.3,0.1."

HAH: "I'd take the two extremes out anyway ... it's fairly ubiquitous ... it's tolerant over a wide range of organic pollution ...I wouldn't expect to find it very common in B1a, and I wouldn't expect to find it present in B4."

WJW: "In terms of B1b,B2,B3?"

HAH: "It might be more abundant in B2 because of the nutrient conditions ..."

WJW: "Is B3 going to be lower than B2?"[4]

HAH: "Not much down."

The effect of abundance on the distributions was also explored. The following extract illustrates the effect of abundance on the "preferred class" of *Chironomus riparius*:

WJW: ... if *Chironomus riparius* is present you're pretty confident that it's a B3 or B4 category, and if it's abundant you're a little more confident that it's a B4?

HAH: Yes, I think so.

For *Lymnaea peregra*, its occurrence in abundance also modifies its associated probability distribution:

WJW: "Can I ask you then what difference it would make then if *Lymnaea peregra* was abundant?"

HAH: "Established at 3, for abundance it's got to be over 50 ... for most of the primary grazers and

[4] A reference to the probability value within the distribution.

scavengers ... you've only got smaller values for the higher carnivores, predators and rarely-occurring species."

WJW: "Okay it's abundant now, you've found 60 of them in your net ... it's still most likely to be a B2 is it?"

HAH: "Yes."

WJW: "Is it more likely that it's a B2, or is it about the same?"

HAH: "It's pushed the B2 up more than the others."

The values of L_s and L_n in **Table 4.3** were used in the odds-likelihoods Bayesian formulation for the early classification systems. There were however problems which arose both from the interpretation of the values supplied by the domain expert and the methodology. Differences between $P(H|e)$ and $P(e|H)$ for the various taxa was either very subtle (e.g. *Simulium ornatum*, *Chironomus riparius*, *Tubifex tubifex*) or sometimes nonexistent (*Gammarus pulex*, *Asellus aquaticus*). This suggested that the distinction between the two modes was either unclear or not meaningful, from the expert's perspective. The expert also admitted finding this mode of elicitation onerous.

Besides the semantic confusion regarding whether $P(H|e)$ or $P(e|H)$ truly represented the expert's verbal expression of the uncertainty estimates there was also the problem of the crude manner in which the expressions were converted to numerical values. These two major difficulties were tackled together in a way that also incorporated information on more than two taxon states.

4.4.4 Improvements to Elicitation of Probability Measures

4.4.4.1 Evolution of sensor model

Benthic invertebrates are useful indicators of river water quality since many exhibit a preferred range of environmental conditions. This association with ranges of river water quality, the idea of saprobic valency, and the inherent uncertainty of interpreting benthic data, were all influential in adopting a probabilistic interpretation of benthic response to changes in water quality. As argued previously, sample data from the river bed can be seen as sensors of river water quality, like the burglar or fire-alarm sensors used by Pearl (1987) to describe the multi-hypothesis formulation of Bayes' theorem.

The adoption of this sensor model evolved during the knowledge elicitation sessions in which the basic domain concepts were acquired. In this respect, the ecological knowledge provides the rationale for the quantitative approach, in that the probabilistic approach seems consistent with the qualitative knowledge. For the biological classification decision problem, the evidence facilitating that decision is provided by a reference set of benthic organisms, selected by the expert, which are deemed to exist in several states (discussed below). For each indicator taxon, discrete probability distributions were obtained which depict the likelihood of occurrence across the adopted range of water qualities, for each state. These estimates of uncertainty are the expert's personal probabilities, emanating from years of experience and knowledge of the ecological requirements of the benthic organisms.

In view of the inadequacies of the initial elicitation, this sensor model of benthic response was seen to offer a coherent framework for eliciting the probability distributions. The process for doing this is described in the following sections.

4.4.4.2 Incorporation of extra taxon states

The previously elicited uncertainty measures refer to only two states of each indicator taxon: presence and absence. However, it was agreed that information on the abundance of taxa within samples should be used by incorporating extra taxa states of Rare (present in very small numbers for that taxon), Established (or Common), Abundant (present in large numbers for that taxon). These three states together represent the single state of Present.

In set notation the relationship between the states can be written

$$\text{Rare} \cup \text{Established} \cup \text{Abundant} = \text{Present} \qquad \textbf{(4.3)}$$

The Absent set is then the complement of {Present}. Membership of each set is "crisp", i.e. the sets correspond to discrete states. To extend the sensor analogy, one can imagine physical probes in the river bed that emit one of four signals: Abundant if the taxon is present in abundance, Established if the taxon is present in significant numbers but less than abundant, Rare if the taxon is present in very small numbers that are not significant, and Absent if the taxon is not detected by the probe. Signals from these probes provide evidence for one or more of the competing hypotheses to varying degrees. Combination of this evidence should yield the overall support for the hypotheses as a ranked order.

As suggested by the domain expert, the numbers that determine membership of each state set may vary from taxon to taxon. When favourable conditions obtain for the sludge worm *Tubifex tubifex*, a pollution-tolerant species, they occur in large numbers. Not so for *Ephemeroptera* or *Plecoptera*, even in their preferred clean water conditions. It was therefore necessary to obtain from the expert the minimum number that determined membership of the states Established and Abundant for each taxon. These numbers are shown in **Table 4.4** on page 98.

4.4.4.3 Preferred Mode of Numerical Estimate

After the initial elicitation it was discovered that the expert was more comfortable with probabilistic information in the form $P(H|e)$, possibly because of the affinity to the saprobic system. For example:

"I can say with more certainty that 'given that it's there [i.e. a certain species], it came from a certain class river', rather than 'in that class river, that [species] would certainly be present'."

Subsequent uncertainty elicitation was in this mode only.

Since the $P(H_i|e)$ are the posterior probabilities sought by the diagnostic process, they must be transformed into $P(e|H)$ for use in Bayesian updating. For some evidence e^j the relationship is given by Bayes' rule in the form:

$$P(H_i|e^j) = P(e^j|H_i)\frac{P(H_i)}{\sum_{i=1}^{n} P(e^j|H_i)P(H_i)} \tag{4.4}$$

Without information to the contrary, the prior probabilities $P(H_i)$ of the n biological classes can be considered equal.[5] This is a reasonable assumption for a small number of classes. Therefore

$$P(H_i|e^j) = \frac{P(e^j|H_i)}{\sum_{k=1}^{n} P(e^j|H_k)} \tag{4.5}$$

Unfortunately the $\{P(H_i|e^j)\}$ cannot be converted to $\{P(e^j|H_i)\}$ without knowing $P(e^j|H_i)$ for at least one H_i. This means that even if a scheme is devised to elicit the $\{P(H_i|e^j)\}$ without requiring the expert to supply actual numbers, some numerical value must be supplied for one of the $\{P(e^j|H_i)\}$ for $j = 1$ to N sets of evidence. The implications of this will become more obvious when the elicitation scheme itself is considered.

4.4.4.4 Development of Elicitation Scheme

Since each taxon can exist in one of $k = 1$ to s states we theoretically need to elicit $n \times s$ conditional probabilities $P(H_i|e^j_k)$ via some means acceptable to the expert, coupled with s conditional probabilities $P(e^j_k|H_i)$ in numerical form for each of the N taxa. It was

[5] In Bayesian theory, this is known as the Principle of Indifference.

considered that this was too onerous a demand to make of the expert, particularly the requirement to identify these values for each $s = 4$ states. Exploratory discussions with the expert revealed that he could envisage uncertainty estimates in some form for two taxon states directly: Present and Abundant. From the relation in (**4.3**), we can write

$$P(e^j_{\text{Present}} \mid H_i) = P(e^j_{\text{Rare}} \cup e^j_{\text{Established}} \cup e^j_{\text{Abundant}} \mid H_i) \qquad \textbf{(4.6)}$$

Since the sets are disjoint we have by rearranging (**4.6**) and assuming $P(e_{\text{Rare}}) \approx$ zero

$$P(e^j_{\text{Established}} \mid H_i) = P(e^j_{\text{Present}} \mid H_i) - P(e^j_{\text{Abundant}} \mid H_i) \qquad \textbf{(4.7)}$$

In fact an arbitrary value of $P(e_{\text{Rare}} | H_i) = 0.01$ was adopted for all hypotheses H_i and evidence e^j. This was felt justified at the time of elicitation since it was felt that the presence of invertebrates in very small numbers should have little impact on the classification.

If the expert can supply $P(e^j_k | H_i)$ for states Present and Abundant, the unknown conditional probability $P(e_{\text{Established}} | H_i)$ can be determined from (**4.7**). Thus the number of states s' for which uncertainty elicitation was required was reduced to two for each taxon. It remained to decide which of the H_i would be used to obtain the numerical value $P(e^j | H_i)$ for each of the two states s'. (Recall that at least one numerical value of the $\{P(e^j | H_i)\}$ is required to allow conversion between the two probability modes). Again this was determined via discussions with the expert.

Most of the benthic taxa in the group, particularly those with a strong indicator value, have a 'preferred' water quality class, i.e. one in which they are more likely to be found. For each taxon therefore, the expert was asked to (a) show the preferred class and (b) to supply numerical values of $P(e^j | H_i)$ for the two states s' (i.e. Present and Abundant). Likewise, the conditional probabilities $P(H_i | e^j)$ were required for Present and Abundant states - Absent and Established values would be calculated as in (**4.7**). Thus, the expert was required to supply $n \times s' = 10$ conditional probability values.

4.4.4.5 Use of graphical pro-forma

Rather than give the required conditional probabilities as verbal expressions a graphical method was devised by which means the expert could show directly the magnitude of the uncertainty estimate. This scheme was introduced because of the dissatisfaction felt with the *ad hoc* method used to convert the verbal expressions to numerical values, and the determination to constrain the requirement for direct numerical elicitation. The expert expressed the opinion that they were easy to use and that he was comfortable with the task. This was helped by the fact that he could complete these forms away from the interview sessions.

Figure 4.1 shows the pro-forma designed for this graphical method. For each taxon e^j the expert supplies the population numbers at which it can be considered Established and Abundant. These numbers are specific for each taxon: for example the level at which pollution-tolerant taxa such as Tubificidae are considered to be Abundant may be quite different for sensitive species such as *Leuctra fusca*. Moreover field data for benthic sample sites often quote the exact numbers of occurring taxa, which for classification purposed would need to be converted into one of the three states comprising "present". This is possible if one knows the threshold levels for the two states Established and Abundant.

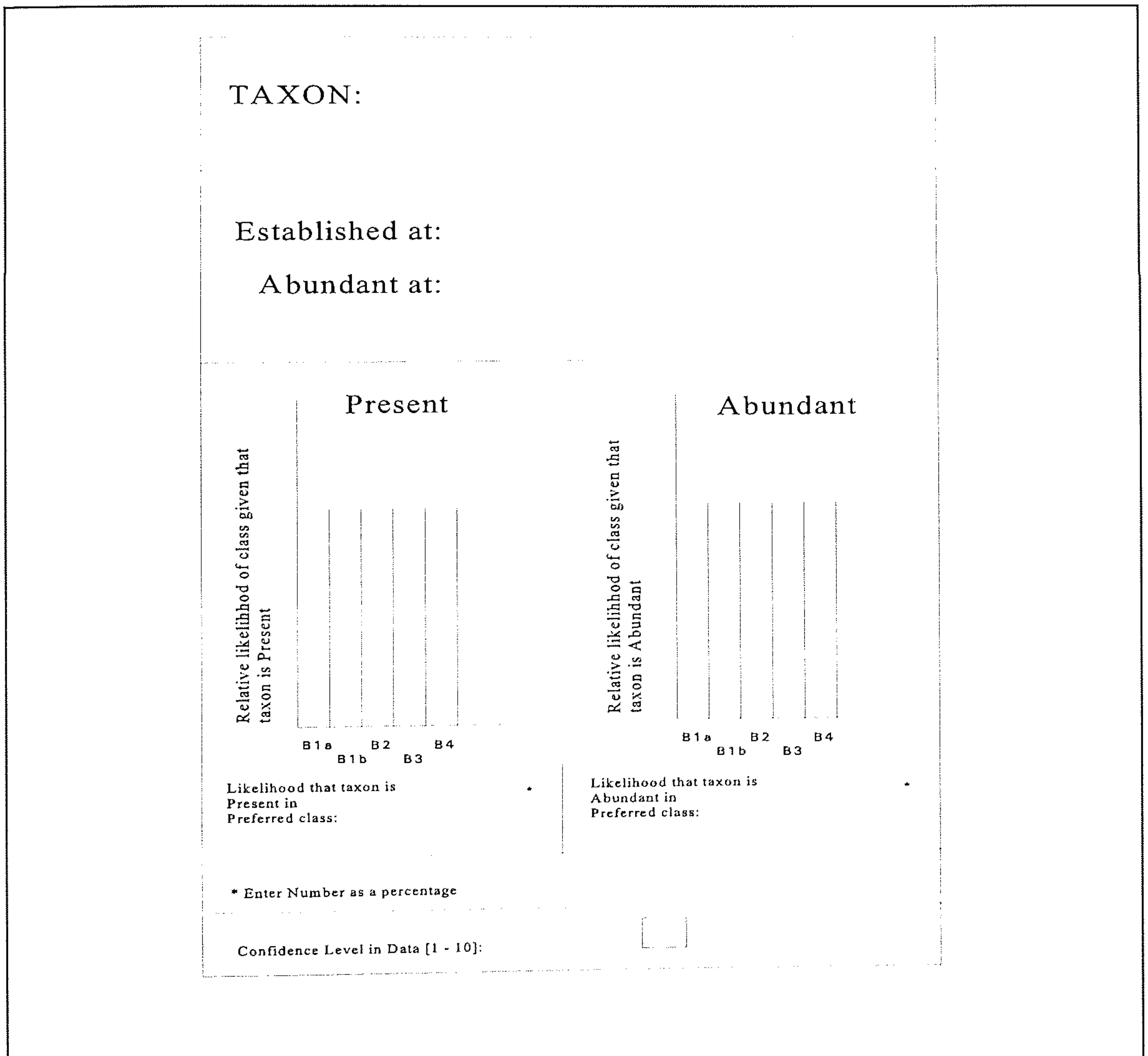

Figure 4.1 Pro-forma used to elicit uncertainty measures for the indicator taxa

The form contains two blank histograms for Present and Abundant, on which the expert was asked to mark a horizontal line within each quality class to form a bar, showing the size of the uncertainty estimate. Rather than constrain or direct the indication via some arbitrary scale, the vertical axis was left without a scale as shown, allowing the expert considerable freedom in deciding the placement of each line within the classes.

The pro-forma refers to 'relative likelihoods' rather than 'probabilities' to reinforce the notion that the main part of the elicitation exercise is pictorial, rather than numerical. Unfortunately maintaining this modality for the entire pro-forma was not possible, since as stated above numerical values for $P(e^j|H_{\text{preferred}})$ were required for two states. To help the expert in this difficult task, the following question was put forward to aid the elicitation of the numbers: "Given that the river water quality is the preferred class for this taxon, what are the chances that it is [Present/Abundant] at this site?", or "If you sampled *n* similar sites, knowing their quality to be the preferred class for this taxon, what percentage of these would yield the taxon [Present/Abundant]?"

To understand the difference between the Present and Abundant figures it may help to consider an example. Consider the sludge worm *Tubifex tubifex* whose preferred class is B4. The value of $P(e_{\text{Present}}|H_{\text{B4}})$ supplied by the expert was 90% (i.e. 0.9). The expert was further asked what proportion of this 90% would find the taxon Abundant, rather than merely Present. It was thought that 75% of these Present samples would in fact contain the taxon in Abundance. The required figure for $P(e_{\text{Abundant}}|H_{\text{B4}})$ is then 0.75 x 0.9 = 0.675. This was repeated for each of the j = 1 to N taxa in the indicator group.

Transformation of the expert's indications of $P(H_i|e^j)$ into numerical values is straightforward. Recall that the denominator of equation **(4.5)** can be found for the two probability modes for the preferred class:

$$\sum_{i=1}^{n} P(e\,|\,H_i) = \frac{P(e\,|\,H_{\text{preferred}})}{P(H_{\text{preferred}}\,|\,e)} \tag{4.8}$$

and can be considered a "scale factor" relating the two probability modes. The denominator of **(4.8)** is found from measuring the column height for the preferred class. The numerator can be derived from the numerical values given by the expert for the likelihoods of finding the taxon in each of the two states *s*'. Once the scale factor is known, **(4.5)** can be rearranged to calculate $\{P(e^j|H_i)\}$ for all *i*. For completeness, these figures are then used to find the numerical values of $P(H_i|e^j)$ corresponding to the column heights.

4.4.4.6 An example of deriving numerical estimates

An example may make this process clearer. From data supplied by the domain expert, the freshwater shrimp *Gammarus pulex* has a preferred biological water quality class of {B1b}, with a probability of being Present in these waters of 0.95. Therefore, this taxon has a high frequency of occurrence. Its probability of being Abundant (occurring in numbers greater than 50) in {B1b} is 0.7 (approximately three-quarters of all the samples in which the taxon is Present in fact contain *Gammarus* in large numbers).

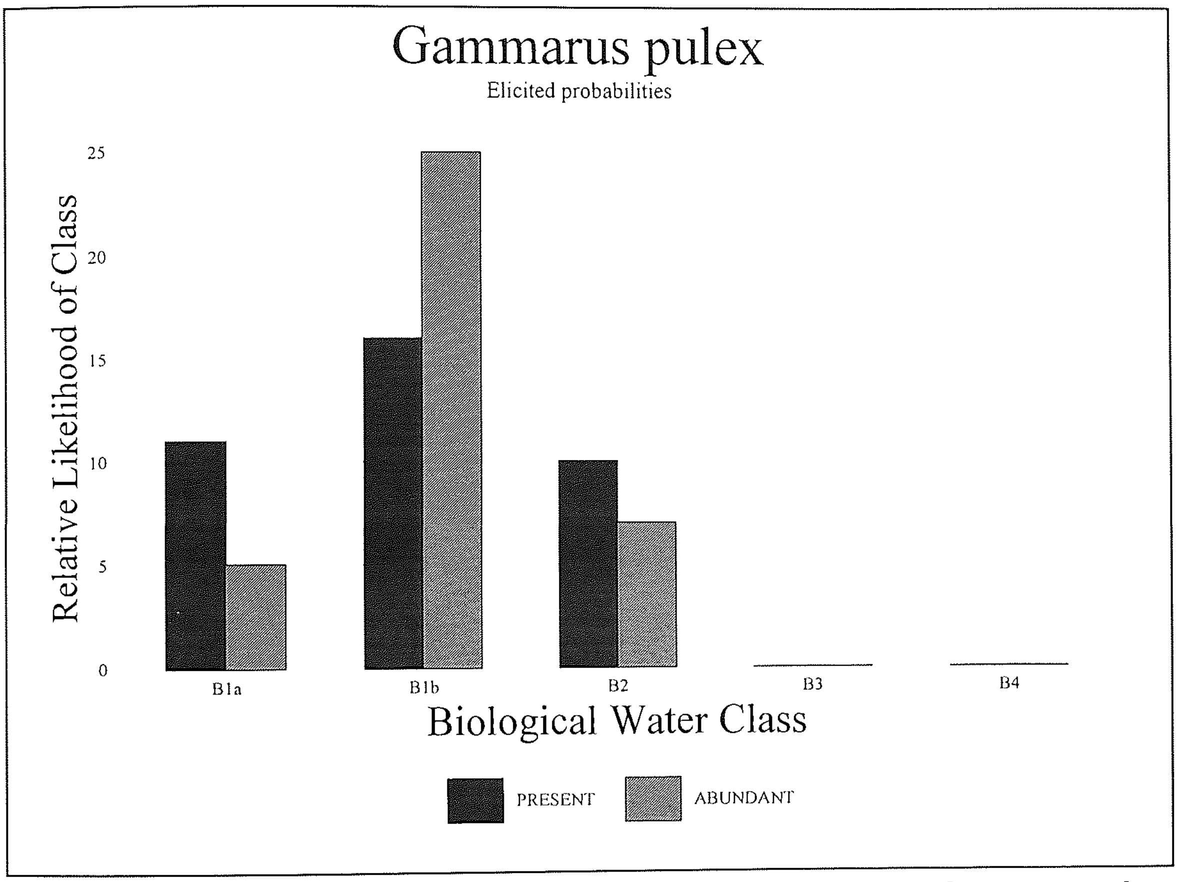

Figure 4.2 Results of graphical elicitation of probability measures for *Gammarus pulex*

The column heights for the $P(H_i|e_{\text{Present}})$ drawn by the expert are {11,16,10,0,0} (arbitrary scale) for the $i = 1$ to 5 classes while those for $P(H_i|e_{\text{Abundant}})$ are {5,25,7,0,0}. These figures are shown in **Figure 4.2**. The significance of abundance for class B1b is apparent.

Equation **(4.8)** is then used to derive the "scale-factor" between the two conditional probability modes. The numerator is the figure specified by the domain expert, while the denominator is equivalent to the column height of the preferred quality class. For the state of Present, the factor is therefore 0.95/16 = 0.06. Thus $P(e_{\text{Present}}|H_{\text{B1a}})$ = Column-height[B1a] x Scale-factor = 11 x 0.06 = 0.66. This is repeated for each H_i. The numerical values corresponding to the histogram indications can then be found from **(4.5)**, where the denominator is the summation of the $P(e_{\text{Present}}|H_i)$.

This sequence of operations is carried out for the histograms for the Abundant state, with the additional requirement of deriving $P(e_{\text{Abundant}}|H_{\text{preferred}})$ from the two numerical likelihood values: 0.95 x 0.75 ≈ 0.7. The discrete conditional probabilities $P(e_k|H_i)$ for all states $k = 1$ to s can then be derived from the $P(e_{\text{Present}}|H_i)$ and $P(e_{\text{Abundant}}|H_i)$. Those for the Established state are derived from **(4.7)**, while the Rare state values were assigned an arbitrary low value. Thus the probability distributions for Rare data are uniform, showing that this evidence is effectively neutral. The $P(e_k|H_i)$ for the state of Absent is then 1 - $P(e_{\text{Present}}|H_i)$.

Figure 4.3 show the derived conditional probabilities $P(H|e)$ for *Gammarus pulex* in the states Present, Established, Abundant and Absent. **Figure 4.4** shows the equivalent for $P(e|H)$. Additionally the likelihood ratios $\{L_s\}$ for each sensor state can be derived for the odds-likelihood formulation of Bayes' rule. Substituting **(4.1)** into **(4.2)** we have

$$L_s = \frac{P(e_k|H_i)P(\neg H_i)}{\sum_{j \neq i} P(e_k|H_j)P(H_j)} \qquad \textbf{(4.9)}$$

for some state k.

All the calculations were automated by entering the information from the graphical pro-forma into a spreadsheet for each taxon in the indicator group. Note that the distribution for *Gammarus pulex* reflects the fact that this taxon is more likely to be abundant in B1b waters due to predation by trout in the higher-quality B1a. Thus, the distributions can partially embody autecological knowledge. According to the expert:

"...the reason why I've got *Gammarus* as "Very Likely" there and "Likely" there isn't that the water quality is worse for *Gammarus* in a grade B1a river [compared] to a grade B1b river. In a grade B1a river, these a mostly upland trout streams - there is a predation on *Gammarus*, which is heavy, by the trout, and therefore you get few *Gammarus* there. The trout are less common in the lower quality river, and therefore the *Gammarus* don't have any competition ... Things rarely occur in abundance under optimum conditions, because you have to take competition into account."

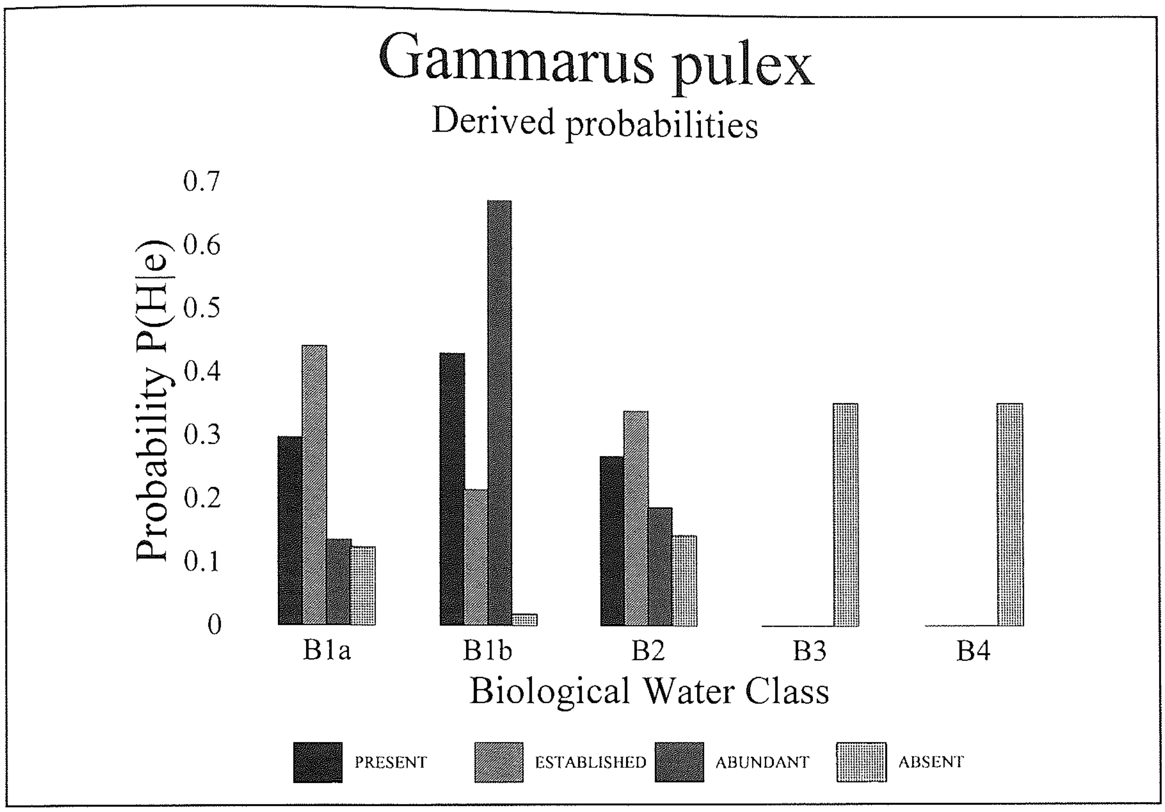

Figure 4.3 Numerical values of P(H|e) for states of *Gammarus pulex* derived from elicited data

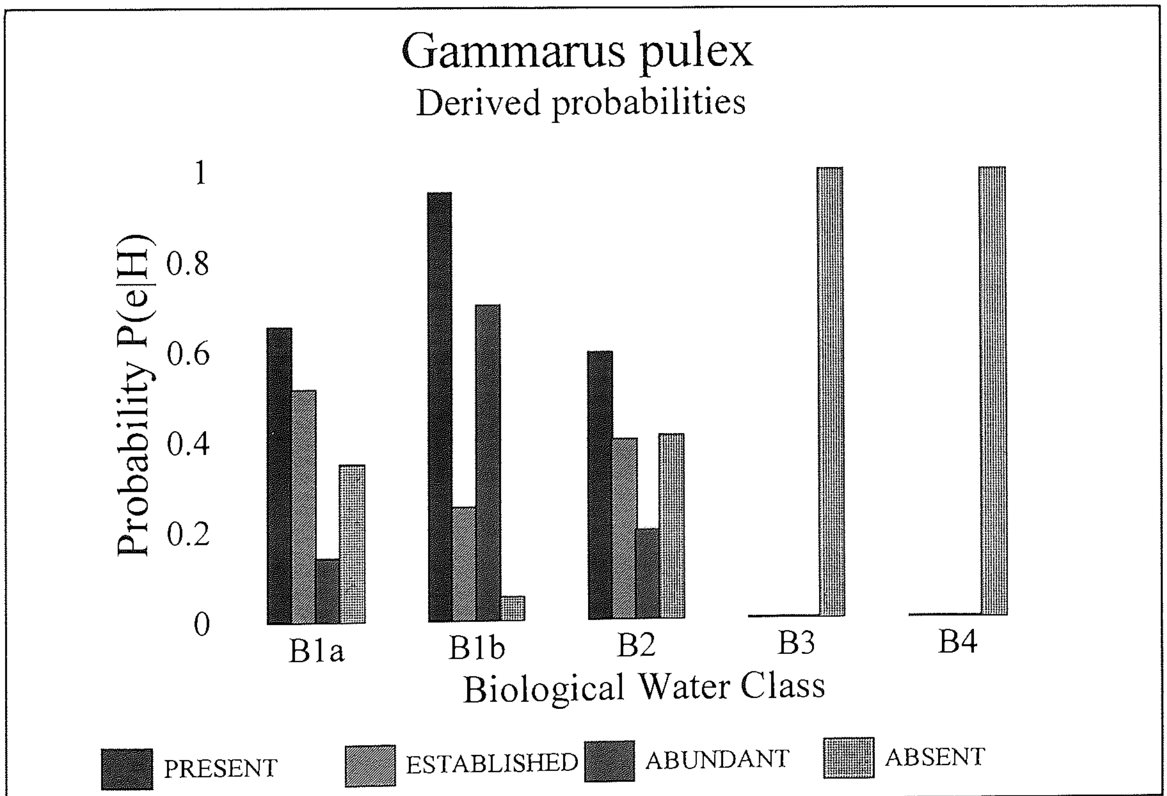

Figure 4.4 Derived numerical values for P(e|H) for several states of *Gammarus pulex*

4.4.4.7 Modification and enlargement of the indicator group

During the knowledge acquisition sessions the size and scope of the indicator group were also considered. Examination of benthic sample data from the Yorkshire Water region[6] showed a large proportion contained taxa that were not regularly identified below family of genus level, which were still considered of high indicator value by the domain expert. Certain species within the initial indicator group of ten fell into this category: sample data often referred to the family TUBIFICIDAE rather than the species *Tubifex tubifex*, while *Leuctra* and *Rhyacophila* were rarely identified below the level of genera. The motivation for initially considering species-level data was the higher level of information so provided; however the available sample data that would form the inputs to the classification system and provide the expert's "reference" classifications meant that a compromise was required, since the elicited probability distributions for the species could not be adjusted for higher taxonomic levels. These three species within the indicator group of ten taxa were therefore replaced by their equivalent genus or family level taxa.

By taking into account the occurrence of the various taxa across a range of sample data, along with their usefulness as quality indicators, a further ten taxa were added to the indicator group after consultation with the expert. These were followed by a further twenty-one to make a total of forty-one taxa in the indictor group, shown in **Table 4.4**. The order of presentation order follows that of Maitland (1977). The numbers at which each taxon is Established and Abundant are given in the table. An indicator taxon occurring in numbers below the threshold value for Established would be considered in the Rare state.

[6] The sample data used for the classification experiments is described in Chapter 5.

Table 4.4 Set of 41 Indicator Benthic Taxa Selected by Expert

Group	Taxon	E	A	Group	Taxon	E	A
C	*Polycelis nigra*	2	10	C	*Ephemerella ignita*	2	20
C	*Dendrocoelum lacteum*	2	10	B	*Caenis spp.*	3	20
C	*Potamopyrgus jenkinsi*	2	50	C	*Amphinemura sulcicollis*	2	10
C	*Bithynia tentaculata*	2	20	A [2]	*Leuctra spp.*	3	20
A	*Lymnaea peregra*	3	50	C	*Isoperla grammatica*	2	10
C	*Planorbis spp.*	2	10	B	HALIPLIDAE	3	20
C	*Ancylus fluviatilis*	2	20	C	DYTISCIDAE	2	10
C	*Sphaerium spp.*	2	20	C	ELMINTHIDAE	2	10
C	*Pisidium spp.*	2	20	B	*Sialis lutaria*	3	10
A [1]	TUBIFICIDAE	3	200	A [3]	*Rhyacophila spp.*	3	20
B	LUMBRICULIDAE	5	100	C	*Glossosoma spp.*	5	50
B	*Glossiphonia spp.*	2	10	C	*Agapetus spp.*	5	50
B	*Helobdella stagnalis*	2	10	C	POLYCENTROPODIDAE	2	20
A	*Erpobdella octoculata*	3	20	A	*Hydropysche angustipennis*	3	50
C	HYDRACARINA	2	20	B	Other HYDROPYSCHIDAE [4]	3	20
A	*Asellus aquaticus*	3	50	C	HYDROPTILIDAE	5	50
A	*Gammarus pulex*	3	50	C	LIMNEPHILIDAE	2	20
B	*Baetis rhodani*	3	50	C	CERATOPOGONIDAE	2	10
B	*Rhithrogena spp.*	3	20	A	*Chironomus riparius*	5	100
C	*Heptagenia spp.*	2	10	A	*Simulium ornatum*	3	50
B	*Ecdyonurus spp.*	3	20				
	E: Established at				A: Abundant at		

Group A - initial group of 10 taxa

Group B - additional 10 taxa

Group C - extra 21 to form 41 overall

1.Originally *Tubifex tubifex*

2.Originally *Leuctra fusca*

3.Originally *Rhyacophila dorsalis*

4. Other than *Hydropsyche angustipennis*

More detailed tables that summarise the graphical data elicited via the pro-forma for the forty-one taxa, and a summary of the derived $P(H_i|e)$, are presented in appendices A2 and A3.

4.4.5 Information gain from Abundance

The information content of the probability distributions can be readily seen by calculating a simple 'indicator value' (or *I*-value) based on the maximum value within the distributions, divided by the number of non-zero (adjacent) histograms. Thus a strongly 'peaked' distribution will have a high indicator value, while a flatter distribution spanning three classes (for instance) will have low value. This is not dissimilar to the indicator value used in the saprobic system. Indicator values were calculated for the states of present and abundant, i.e. directly from the graphically elicited distributions. For most of the indicator

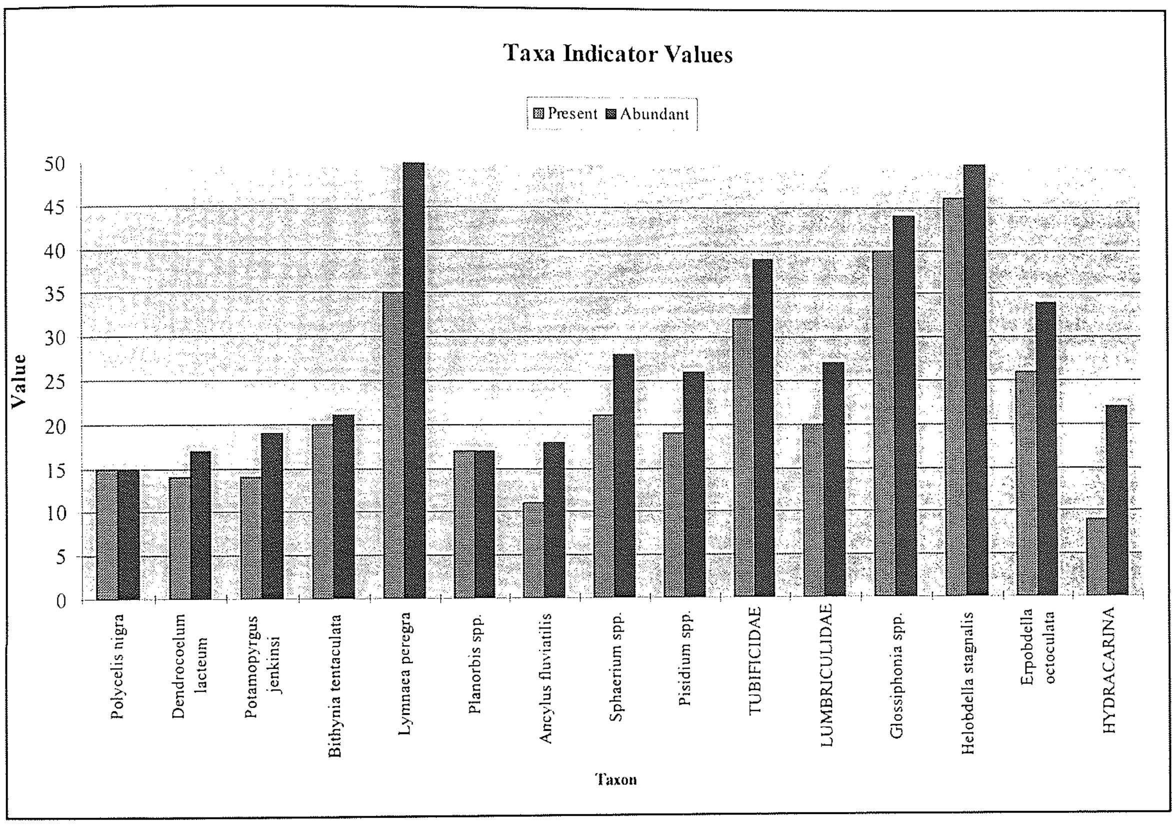

Figure 4.5 Indicator values of probability distributions for states of present and abundant. (A subset of the indicator group is shown).

group, there is a clear information gain, as measured by the *I*-value, by taking account of abundance, as depicted in **Figure 4.5**.

4.4.6 Value of elicited knowledge

The specificity and detail of this knowledge are considered a valuable resource, in that it is believed that an exercise of this type has not been conducted before for benthic invertebrates. The probability distributions were prerequisite for the classification experiments described later in this thesis and for a related project in biological classification using neural networks (Ruck, 1995). The resulting graphical method could be used for other invertebrates, or alternatively to calibrate two or more experts' distributions for particular taxa to arrive at "consensus" distributions. (This is discussed in more detail in Chapter 8).

Qualitative and quantitative knowledge of particular benthic invertebrates was elicited from detailed interviews with an expert in the field of biological surveillance, who was both cooperative and fully committed to the aims of the project.

4.5 Supplementary Work

4.5.1 Classification of sample data

Benthic data from river surveillance was classified by the domain expert during the knowledge elicitation sessions. This expert classification is described in Chapter 5. Prototype classification programs were developed to assess the efficacy of the elicitation method for obtaining the conditional probabilities and the viability of the probabilistic approach. This work is described in Chapter 6. However some early results suggested that use of the directly elicited distributions produced incorrect results under certain conditions. This problem is considered in the next section.

4.5.2 Adjustment of derived conditional figures

The probability values corresponding to the conditional probabilities were derived by measuring the heights of the histograms. Examination of **Figure 4.2** suggests that, according to this distribution, there is zero likelihood of detecting *Gammarus pulex* in classes B3 and B4 in states Present or Abundant. While this distribution may reflect the expert's opinion regarding this organism for riffles, it was felt that it may "unfairly" rule out the hypotheses B3 and B4 as impossible, and that some redistribution of the probability mass may be required. This conjecture arose after using the distributions from the first ten indicator taxa in prototype Bayesian classification software. In combining evidence, hypotheses that had zero support from any one sensor were immediately vetoed, thereby discounting from consideration support from other sensors. Remaining support was then often highly focused on one or two quality classes, in a way that did not accord with the diversity of sensor data provided by the sample.

As discussed in Chapter 3 this phenomenon occurs when combining evidence that is both "crisp" and highly contentious. The difficulty of dealing with conflicting evidence is discussed in Chapter 6, but even before detailed experimental work using the uncertainty estimates it was felt that the elicited distributions may have been too crisp.

Thus the domain expert may have exerted the psychological bias suggested by Tversky and

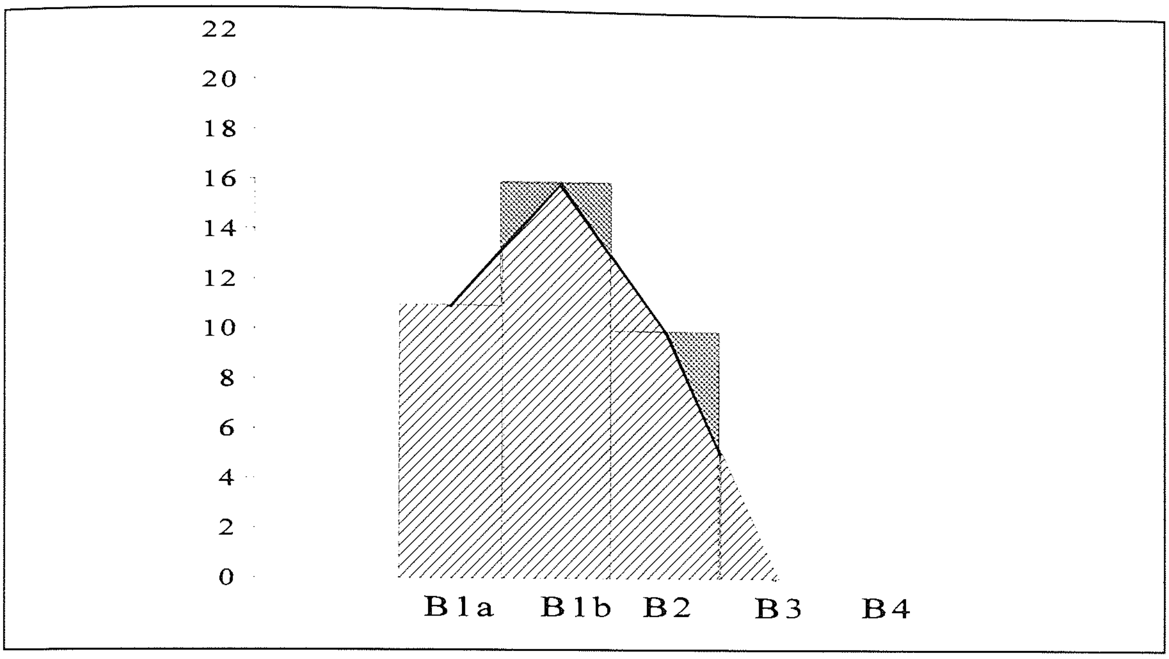

Figure 4.6 Construction superimposed on elicited distributions

Kahneman (1974) in which highly unlikely events were ruled as completely impossible, rather than retain a non-zero probability. The discrete probability distributions based on the heights of the histograms were therefore adjusted to reduce their crispness, so that probability mass was redistributed across the five competing classes.

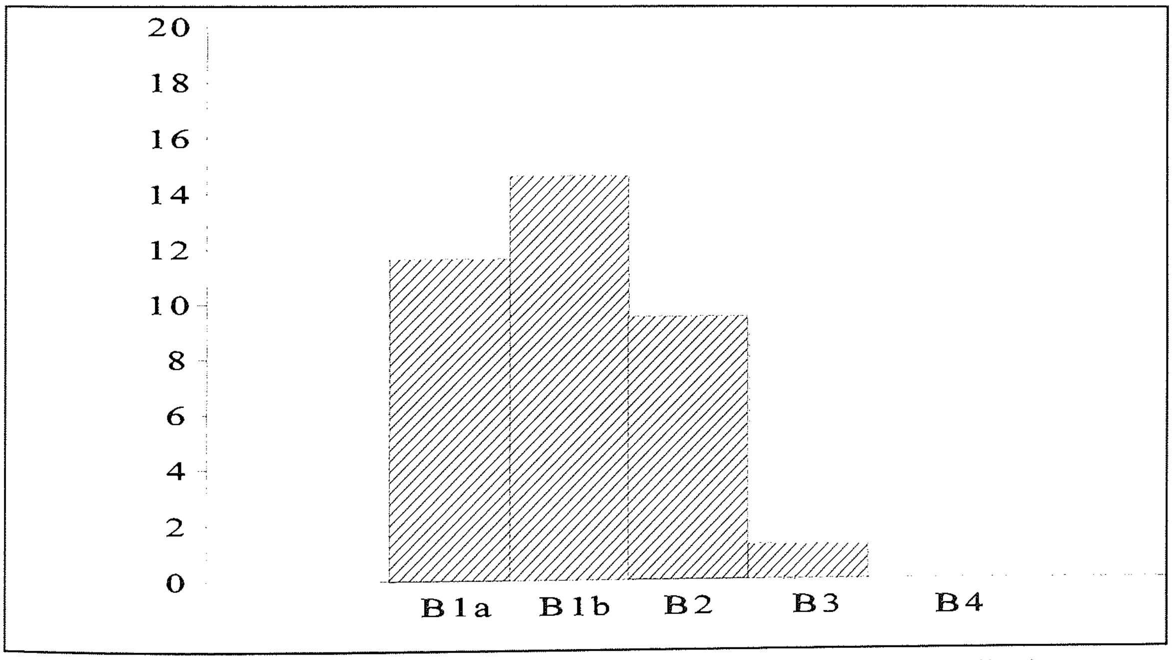

Figure 4.7 Adjusted histogram with heights equal to areas of elicited distributions

This redistribution was carried out using an *ad hoc* graphical method, in which the $P(H_i|e)$ correspond to the mass areas within each quality class under a line connecting the midpoint of each histogram bar (**Figures 4.6** and **4.7**). Each bar is considered of unit width. If P_i is the probability measure derived from measuring the heights of the histograms provided by the expert for hypotheses H_i then the measures $P^A{}_i$ corresponding to the areas are:

$$
\begin{aligned}
P^A{}_i &= 0.875P_i + 0.125P_{i+1} && i=1 \\
&= 0.125P_{i-1} + 0.75\ P_i + 0.125P_{i+1} && i=2,3,4 \\
&= 0.125P_{i-1} + 0.875P_i && i=5
\end{aligned}
\qquad \textbf{(4.10)}
$$

These formulae follow from the graphical construction used. The performance of the classification systems using probability measures based on the "raw data" provided by the domain expert could then be compared with the adjusted measures using this construction.

4.6 Summary

This chapter has discussed the principles of knowledge acquisition and elicitation for expert systems, and particularly examined the methods for eliciting measures of uncertainty. The practical work of knowledge acquisition from an acknowledged expert in biological surveillance has been described. Elicitation of domain concepts and qualitative knowledge on benthic invertebrates was carried out over several interviews with the expert. A model of benthic sensors of biological quality was developed from this qualitative knowledge, of indicator taxa existing in discrete states of abundance, including absence. In this model, observation of indicator taxa in these states leads to probabilistic evidence supporting (or refuting) the biological classes adopted for this decision problem.

A group of benthic invertebrates was selected by the expert during the knowledge acquisition sessions, which were deemed useful indicators of water quality over the five biological classes. The methods by which numerical conditional probability distributions for the indicator taxa were elicited were described, commencing with verbal expressions of the likelihood of water quality when an indicator is present and the converse. Development of a novel graphical method was described in which the distributions for four states of abundance were incorporated, with a minimal requirement for the expert to specify numerical values. The method was used to elicit probabilistic knowledge of organic pollution in riffles as indicated by forty-one benthic taxa. Threshold levels of abundance specific to each indicator were also elicited and recorded.

This domain knowledge is a valuable resource, and the manner in which is was elicited is believed to be unique in this field.

Chapter 5

Expert Classification of Benthic Data

5.1 Introduction

This chapter has two main purposes: one to describe the construction of the benthic sample data sets, and the second to record the expert's direct interpretation of that sample data. Both activities required the close cooperation of the domain expert, who provided access to the invertebrate data, selected the groups of indicator taxa, and gave his considered opinion on river water quality where necessary. The expert's direct interpretation of sample data underpins the adoption of a discrete biological classification system for river water quality, since he was called upon to interpret the data with respect to these classes.

The chapter begins with a description of the preliminary data set and the 'discovery' of the expert's use of intermediate classifications. The facility with which the expert classified restricted sample data gave confidence that the biological classification of river water quality was a natural mental task. The identification of forty-one key indicator taxa, and the elicitation of their conditional probabilities for river water quality (described in Chapter 4), was the starting point for the construction of the Yorkshire Data set described below. These taxa form a 'reference set' against which the expert's classifications and the knowledge-based classifications can be made. The strengths and deficiencies of the data set are evaluated. Several biotic indices are calculated and compared with the biological classifications. It is suggested that, if the expert's classifications are taken as a reference, the large overlap in ranges makes distinguishing gradations in river water quality difficult if these indices are used.

The difficulties in utilising the NRA Severn-Trent data are discussed. Finally, the summary highlights the value of the Yorkshire data and the classification of benthic data by an acknowledged expert.

5.2 Preliminary data

5.2.1 Origin

This data was gathered as part of the knowledge acquisition exercise and to test prototype classification systems, described in Chapter 6. While the size of the set is too small to allow any significant conclusions to be drawn from the classification experiments, its importance lies in its use to observe the expert's use of biological classification, the elicitation of domain knowledge on recovery from organic loading in a river, and the identification of intermediate classes.

Table 5.1 shows the five samples used to construct the set. The samples were presented by the expert for discussion at one of the knowledge acquisition interviews, as representative of 'typical' benthic samples, and illustrative of important domain ideas. For instance the samples called 'PS 3', 'PS 4' and 'PS 5' correspond respectively to samples taken from stations located upstream of an organic effluent, downstream and further downstream. These last three samples show the principle of self-purification that takes place in a river in response to organic pollution, a model that is the basis of the saprobic system. The sample called 'PS 4' represents a benthic community that has displaced the type of community that existed upstream of the discharge, being able to take advantage of the increased nutrients. In 'PS 5', the community has further recovered as purification proceeds.

5.2.2 Classification of preliminary data

The presentation of these samples coincided with the choice of the initial group of ten indicator taxa for which the expert also supplied discrete probability distributions as "verbal expressions" of likelihood, discussed in Chapter 4 on Knowledge Elicitation. This probabilistic knowledge was used to construct simple knowledge bases based on these ten taxa. To test these 'automatic classifiers' against the expert's classification, the preliminary data set was constructed from the actual samples in **Table 5.1** in terms of the ten indicator taxa. The expert was then presented with the restricted list of sample data and asked to classify the samples with respect to the discrete biological classes. At this stage the expert was not told of the identity or source of the restricted samples, although with only five samples this was hardly an imposition. More onerous was the omission of the full sample data. The justification for this restriction was to ensure that both the reasoning systems and the expert were using the same information to classify a sample.

Table 5.1 Benthic samples used to construct preliminary data set

PS 1	
Tubifex tubifex	c. 11000
Helobdella stagnalis	1
Erpobdella octoculata	7
Asellus aquaticus	20
Sialis lutaria	1
Chironomus riparius	27

PS 2	
Ancylus fluviatilis	8
Bithynia tentaculata	1
Lymnaea peregra	10
Tubifex tubifex	909
Erpobdella octoculata	96
Helobdella stagnalis	1
Asellus aquaticus	582
Haliplidae	1
Hydropysche angustipennis	4
Caenis moesta	11
Simulium ornatum	4
Chironomus riparius	180

PS 3[1]	
Polycelis cornuta	7
Lumbriculidae	19
Piscicola geometra	1
Gammarus pulex	55
Corophium curvispinum	14
Baetis rhodani	32
Rhitrogena semicolorata	27
Ecdyonurus venosus	15
Ephemeralla ignita	12
Ephemera danica	12
Brachytera risi	8
Amphinemura sulcicollis	21
Capnia bifrons	17
Leuctra fusca	2
Chloroperla torrentium	1
Rhyacophila dorsalis	4
Glossosoma boltoni	58
Polycentropus flavomaculatus	3
Hydropsyche fulvipes	19

PS 4[1]	
Tubifex tubifex	1049
Lumbriculidae	13
Helobdella stagnalis	1
Erpobdella octoculata	1
Erpobdella testacea	2
Asellus aquaticus	3
Baetis rhodani	2
Chironomus riparius	187
Prodiamesa olivacea	29
Brillia longifurca	47
Lymnaea peregra	2

PS 5[1]	
Dendrocoelum lacteum	5
Tubifex tubifex	62
Lumbriculidae	71
Glossiphonia complanata	4
Helobdella stagnalis	45
Erpobdella octoculata	14
Erpobdella testacea	3
Gammarus pulex	2
Asellus aquaticus	345
Baetis rhodani	16
Hydropysche angustipennis	241
Chironomus riparius	4
Prodiamesa olivacea	28
Brillia longifurca	29
Simulium ornatum	82
Potamopyrgus jenkinsi	2
Lymnaea peregra	11
Physa fontinalis	8
Ancylus fluviatilis	2

Average numbers per 0.5 min heel sample

1. Source: Manual on Biological Surveillance using Benthic Macro-invertebrates (Hawkes, 1977).

The preliminary data set constructed from the five actual samples is presented in **Table 5.2**. Samples Prelim 1 to Prelim 5 correspond to the actual samples PS 1 to PS 5 respectively, but omit reference to taxa not in the indicator group. Recorded underneath each preliminary sample is the expert's classification based on this information. In three cases the domain expert, acting on his own initiative, used intermediate classifications. Originally, he used plus signs (+ or ++) to show a worsening of water quality from the 'base class', minus (- or --) an improvement.[1] According to the expert, '++' denoted 'towards upper limits - inferior', while '+' was 'somewhat inferior' to the base class that it qualified. Later, in classifying Severn-Trent NRA benthic data for an associated PhD project (Ruck, 1995) the expert used a 'reverse scale', in which (+) denotes *improving* quality, (-) worsening quality. With the original scale, the expert occasionally used double increments (++ or --), for the reverse scale only single increments were used.

To avoid confusion, and to enforce consistency with this associated later work, classifications of all the benthic data used in this project are expressed using this 'reverse scale', although the expert's use of double increments is maintained. The interpretation of the expert's intermediate classifications is an important issue, and is discussed at some

Table 5.2 Preliminary data set with respect to ten indicator taxa

	Prelim 1	Prelim 2	Prelim 3	Prelim 4	Prelim 5
Lymnaea peregra		10		2	2
Tubifex tubifex [1]	11000	909		1049	62
Erpobdella octoculata	7	96		1	14
Asellus aquaticus	20	582		3	345
Gammarus pulex			55		2
Leuctra fusca			2		
Rhyacophila dorsalis [2]			4		
Hydropysche angustipennis		4	19		241
Chironomus riparius	27	180		187	4
Simulium ornatum		4			82
Expert's Assessment [3]	B3--	B3	B1b	B3-	B3+

1. Later changed to Tubificidae. 2. Later changed to *Rhyacophila spp.*
3. Using 'reverse scale' - see text for explanation.

[1] This would be consistent with the saprobic system in which increasing indices reflect worsening quality. The issue of scaling is further discussed in Chapter 6.

length in Chapter 6.

5.2.3 Conclusions: Preliminary data classification

These early classifications showed that the expert was 'comfortable' with the idea of biological classification, adapting it to suit his perception of gradations of river water quality within classes, and could classify samples based on a restricted list of indicator taxa. Results from the prototype automatic classifiers (discussed in Chapter 6) suggested that the probabilistic knowledge obtained from the expert and its use in Bayesian reasoning was broadly consistent with the expert's classifications. This gave confidence in the basic method of eliciting the distributions, and the principle of using 'key' indicator taxa to classify river water quality. As explained in Chapter 4, the knowledge elicitation phase of the project was extended to improve the methodology of eliciting the distributions via the use of graphical pro-forma, and the group of indicator taxa increased in size to forty-one. This took place with the availability of benthic data from the Yorkshire Water region, which formed the basis of the classification experiments described in Chapters 6 and 7.

5.3 Yorkshire Water data set

5.3.1 Origin

Obtaining a larger set of benthic data was clearly important, first to exercise the expert's ability to classify samples of which he had no prior knowledge, and secondly to conduct a more realistic programme of computational experiments in which decision algorithms based on methods of uncertain reasoning could be tested on their ability to emulate the expert's biological classification. This section describes the work done to construct this data base, and the descriptive analysis of the sample data.

The author obtained a report by Scientific Services of the Yorkshire Water Authority (YWA) entitled "Collation of Invertebrate Distribution Records 1971-76" (Yorkshire Water Authority, 1976). Despite its age, the report contained one of the most comprehensive surveys of benthic invertebrates for British waters that was available to the author for the work described in this thesis. According to its preface:

> "A large volume of biological data is collected by Y.W.A. Head Office as a result of routine monitoring and special investigations. Biological monitoring of water quality is becoming increasingly important and the use of water quality indices based on invertebrate collections is undergoing a major international review. The understanding of the communities found in waters of various types is vital to this review ..."

Data on invertebrate distribution was collected by the YWA and its predecessors since

1956. The data from 1971-76 surveys, and all special surveys since Spring 1971, were collated by the Authority for a total number of sites of about 1021, including rivers, canals and still-water sites. Many were sampled at irregular intervals, or once only, and the presentation of the results of every sample would have produced a very large collation. Therefore, within the report the data was presented as a single list of taxa for each site with a note of the number of samples taken. Details of particular samples were also obtained by the author by visiting the YWA's Head Office in Leeds.

5.3.2 YWA Sampling and Recording Methods

The vast majority of data to which the report refers were collected by qualitative means, using an FBA (Freshwater Biological Association) pond net. Twenty standardised 'kicks' or the equivalents in 'sweeps' were generally used in rivers and ponds, each kick disturbing an area about 30 x 20 cm. Some smaller streams were sampled with 10 'kicks' or equivalent. At the time the report was written, data was still being collected on the relative efficiency of these methods. Some apparent temporal changes were due to sampling inconsistencies. These problems serve to highlight the information loss and inherent uncertainty present in biological sampling programmes.

Although numbers of occurring individuals were recorded for each sample, these were 'translated' into a numerical abundance scale for the combined faunal list by the compilers of the report. The scale was constructed by randomly selecting 100 samples, and treating all the 800 taxa found therein as identical. The cumulative frequency of taxa against number of individuals was calculated and plotted, this being a measure of the number of times taxa were recorded as 1 individual, 1-2 individuals, 1-3 individuals, etc. irrespective of the identity of each taxon. The plot was then divided into quartiles, with the top quartile subdivided, resulting in five divisions. These five divisions of cumulative frequency corresponded to an abundance scale 1-5, in which '1' represented 1 individual per sample, up to '5' representing >200 individuals per sample.

The DoE (Department of the Environment) classifications were given in many cases, on a 10-point scale from A through A/B, B/A, etc. to D. The range of values was given where several samples were taken, e.g. A-B, B-D etc. Interestingly, there were several instances where a surveyor of a particular site had recorded his or her assessment of biological quality using a system equivalent to the NWC classes, i.e. that adopted in this work. Unfortunately, these classifications were not included in the collation.

The compilers recorded their own opinions on the usefulness of various benthic groups for water quality assessment. The following taxa were found in all rivers and were deemed to have no value in such analyses: Turbellaria, Sphaeriidae, Oligochaeta, *Baetis rhodani*, Dytiscidae, Elmidae, Chironomidae, 'larger Diptera' and Ceratopogonidae. However, at least six species of Oligochaeta were found in heavily polluted (class D) waters, suggesting that this group "could provide useful quality information at the lower end of the range". Plecoptera were "basically confined to the upper reaches of the rivers and their value in water quality estimation is, we feel, overrated" (Yorkshire Water Authority, 1976).

5.3.3 Presentation of taxonomic lists

The order of presentation in the collation is as 'Limnofauna Europaea' (Illies, 1967), which at the time was the most recent work covering all aquatic fauna. **Table 5.3** shows part of the taxonomic listing. Altogether there are 390 entries in the list, in taxonomic order, allowing the collation to record benthic data at whatever level of identification was achieved at sampling. For instance, identification to species level of *Dendrocoelum lacteum* would result in an occurrence being recorded against this entry; higher-level identification may have resulted in the taxon being recorded as an occurrence of Turbellaria.

Sampled sites in the Yorkshire Area were assigned unique numbers by the Authority and listed in order of major rivers and their tributaries. Since the total number of listed sites was considerable, a subset was chosen from the rivers Aire and Calder and their tributaries, in the Western and South-Western Divisions of the region. Sites listed for other rivers such as the rivers Don and Rother were found to be generally unproductive or were declared to be badly polluted by toxic materials. These were therefore excluded. However, the chosen sites within the Aire and Calder catchments are representative of the range of DoE classifications. **Figure 5.1** shows a map of the Calder catchment, taken from recent information compiled by the NRA (Northumbria and Yorkshire region) on biological water quality. The map makes reference to biological classes B1a to B4, still in use at the time by the authority and its successor, the Environment Agency.[2]

---

[2] This data is from Viki Hirst, an ecologist with the Environment Agency. In an informal trial to investigate the reliability of biological classification in the Northumbria and Yorkshire region, classifications of samples by different ecologists were highly correlated (V. Hirst, personal communication).

Aston University

Content has been removed for copyright reasons

Figure 5.1 Map of Calder catchment, Yorkshire region. (Biological classification of sites is not shown). Several sites from this catchment were selected for the sample database. (Source: Environment Agency (Northumbria and Yorkshire Region)).

5.3.4 Filtering of site data

In order to regularise the taxonomic data for each site, and in preparation for the classification experiments, the lists were transformed into occurrences of the group of forty-one indicator taxa selected in conjunction with the domain expert, for which probability distributions had been elicited. The decision was therefore made, as with the preliminary data, that both the expert and the classifiers would be classifying sites in response to the same information. This was felt necessary to allow a meaningful evaluation of classifier performance: both expert and classifier would classify with respect to the 'reference' set of indicator taxa. However, the quality of the both the original and collated sample data was such that some filtering was necessary. Clearly, the closer the correspondence between the report data and the transformed data, the better.

The range of the indicator taxa proved to be reasonably comprehensive, in that relatively little distortion of the collated site data was required, with the exception of removing obscure species. Some assumptions had to be made in cases where the 'nearest' associated indicator was at a taxonomic level lower than the site taxon, identified perhaps to order or family level. In several instances references to 'Oligochaeta' were taken as equivalent to Tubificidae, a member of the indicator group. These decisions were made on a site-by-site basis.

Table 5.4 shows an example of taxonomic data for one River Aire site. The first list (a) gives the occurrences of all sampled taxa at this site for surveys conducted between 1971 and 1976. This list has been transformed into (b), a table of abundance levels for taxa which are members of the elicited forty-one indicator group. The suffixes 'A','E' and 'R' refer to 'Abundant', 'Established' and 'Rare', the abundance levels adopted for the knowledge elicitation exercise described in Chapter 4. Translation from the report's abundance levels to those used for the expert's classification took place according to the following heuristics, which incorporate the individual threshold values for the indicator taxa:

if YWA-level = 5 then State = Abundant
if YWA-level = 4 and Abundant-threshold >= 50 then State = Established
if YWA-level = 4 and Abundant-threshold <= 20 then State = Abundant
if YWA-level = 3 and Abundant-threshold >= 20 then State = Established
if YWA-level = 3 and Abundant-threshold <= 10 then State = Abundant
if YWA-level = 2 and Abundant-threshold > 20 then State = Rare
if YWA-level = 2 and Abundant-threshold <= 20 then State = Established
if YWA-level = 1 then State = Rare

Table 5.4 Format of collated Yorkshire Region site taxonomic data

(a) Extract from collated report

TURBELLARIA	2
Potamopyrgus jenkinsi (Smith)	4
Lymnaea (Radix) pereger (Mull.)	5
Planorbis	3
Planorbis planorbis (L.)	+
Planorbis (Bathyophalus) contortus (L.)	+
Planorbis (Gyraulus) albus (Mull.)	+
Ancylus fluviatilis (Mull.)	4
Sphaeriidae	4
OLIGOCHAETA	4
Glossiphonia complanata (L.)	3
Glossiphonia heteroclita (L.)	2
Erpobdella octoculata (L.)	4
Asellus aquaticus L.	4
Baetis rhodani Pict.	2
Elmidae	1
Chironomidae	3
	17

(b) Filtered data for classification

Polycelis nigra	E
Potamopyrgus jenkinsi	E
Lymnaea peregra	A
Planorbis spp.	E
Ancylus fluviatilis	A
Sphaerium spp.	A
Tubificidae	E
Glossiphonia spp.	A
Erpobdella octoculata	A
Asellus aquaticus	E
Baetis rhodani	E
Elminthidae	R
Chironomus riparius	E
	13

River Aire, below Gill Beck. SE 188 395. Site 435F
Number of Samples: 2. DoE Index C
Yorkshire Collation report page: 28.
See text for explanation of abundance levels and filtering.

'YWA-level' refers to the abundance-level used in the cited report (Yorkshire Water Authority, 1976). In the example shown in **Table 5.4**, the reference to 'TURBELLARIA' in list (a) was interpreted in (b) as *Polycelis nigra*. This was determined by inspection of adjacent sites in the collated report, in which occurrences of taxa at an increased level of identification referred to *P. nigra* rather than *Dendrocoelum lacteum*. In agreement with the domain expert, references to Sphaeriidae were accepted as *Sphaerium* spp., Chironomidae as *Chironomus riparius*, and Elmidae as Elminthidae for all sites.

5.3.5 Comparison with original sample data

Some time after the collated data had been classified by the expert, the author gained access to the original sample data for several sites within the Aire and Calder catchments to see how the combined faunal list were drawn up. The collated report refers to more than one sample for the majority of the sites; the combined list therefore represents an amalgam of individual sample data. The samples were hand-written records on 'River Survey results sheets', in which the surveyor recorded physical details of the site, complete with a sketch, and a species list. Physical details, which included the composition of the substratum, confirmed that these were riffle sites. The rigour with which species lists were completed varies considerably, with some sheets showing actual numbers of individuals, while for others the numbers are omitted. Taxonomic levels and rigour of identification also varied: some samples referred to species and to 'Stoneflies' or 'Chironomid larvae' (for instance) in the same list.

These observations, coupled with the fact that the original hand-written records are now somewhat inaccessible, suggests that the collated report provides a more reliable source of data than the original samples. One advantage of filtering the data from the collated report is that the 'information loss' is contained within the process of data preparation; the expert is presented with a list of sample data that is regular and consistent in form. The disadvantage is that the expert is 'removed' from the actual sample data as recorded. As explained this was unavoidable for chronological and logistical reasons. The sample lists extracted from the collated report represented the best available benthic data at the time.

5.3.6 Classification of Sample Data

The fifty sites extracted from the collated report were reduced to forty-two by eliminating those which were (a) an amalgam of a high number of individual samples and (b) those for which several taxa which were not in the indicator group had to be removed in the process of data-filtering. The remaining 'filtered samples' were then deemed to be representative of the site data for a range of riffle sites in the Aire and Calder catchments. The domain expert supplied eleven additional samples from the Midlands area which were transformed into filtered lists to form a database of fifty-three sites.[3] Those sites within this database which were extracted from the collated report are presented in **Table 5.5**. The DoE biological class (Department of the Environment, 1972) can be used to give a broad indication of agreement between the classifications of the expert and those of YWA biologists. This is discussed below.

[3] This database is subsequently referred to as the 'Yorkshire Water data set', notwithstanding the additional Midland sites.

Table 5.5 Aire and Calder sites within Yorkshire Water Data set

River	Site No.	Site Details	Map ref.	DoE Class
Aire	435E	Above eff. channel	SE 187 388	C
Aire	435F	Below Gill Beck	SE 188 395	C
Aire	436A	Above Gill Beck	SE 187 396	C
Aire	438B	Tarn	SE 169 393	C
Aire	438C	Charlestown	SE 158 386	C
Aire	440A	Below Bingley SW	SE 126 383	C
Aire	443A	Crossflats	SE 095 404	B/C
Aire	443C	Below R. Worth	SE 077 422	B
Aire	448A	Cononley	SD 995 469	B
Calder	494	Halifax	SE 098 219	D/C
Calder	495B	Sowerby Bridge	SE 059 235	C
Calder Tributary	545F	R. Holme	SE 152 104	B/C
Calder Tributary	547B	R. Holme Victoria	SE 135 078	B
Calder Tributary	553B	R. Ribble above Washpitt Mill	SE 141 064	A
Calder Tributary	564	Black Brook	SE 099 214	D
Calder Tributary	566A	Black Brook T.P.T.works	SE 069 189	C/D
Calder Tributary	566B	Black Brook	SE 060 176	B
Calder	576A	?		A
Calder	576B	?		A/B
Aire Tributary	596A	Wyke Beck	SE 337 320	D
Aire Tributary	598	Low Beck	SE 284 320	D
Aire Tributary	599	Wortley Beck	SE 262 319	C/D
Aire Tributary	601A	Millshaw Beck	SE 279 317	D
Aire Tributary	601B	Millshaw Beck	SE 278 299	D
Aire Tributary	608A	Bradford Beck	SE 151 376	D
Aire Tributary	608B	Bradford Beck - Hollin Close	SE 153 365	D/C
Aire Tributary	610A	Red Beck	SE 144 363	A
Aire Tributary	611	Clayton Beck	SE 109 323	A
Aire Tributary	613	Loadpit Beck	SE 131 386	A
Aire Tributary	615C	Harden Beck	SE 088 378	B
Aire Tributary	617TB	Ellar Car Beck	SE 069 374	D
Aire Tributary	618	Morten Beck	SE 100 409	A
Aire Tributary	619B	R. Worth Below North Beck	SE 063 408	C
Aire Tributary	620A	R. Worth Hey Bridge	SE 060 402	C
Aire Tributary	620C	R. Worth Below S.W.	SE 045 383	C
Aire Tributary	620D	R. Worth Mytholmes	SE 039 382	B/C
Aire Tributary	621A	R. Worth Providence Lane	SE 034 380	A
Aire Tributary	624B	R. Worth Haworth	SE 035 369	B
Aire Tributary	625	R. Worth Oxenhope	SE 032 354	B
Aire Tributary	634	Long Dam	SD 984 508	A
Aire Tributary	636A	Eller Gill Farm	SD 990 535	A
Aire Tributary	640	Embsay Beck	SE 006 533	B/C
Aire Tributary	624A	Bridgehouse Beck	SE 037 380	C

Sample data was presented to the expert in the form of **Table 5.4**(b) for each site. The expert was asked to classify each sample in terms of the discrete biological classes, on the assumption that (a) the samples were taken from riffles and (b) the type of pollution to which these sites were subject was organic. He was not shown the DoE classifications for the Aire and Calder sites.

Given the number of samples, the expert chose to carry out this exercise away from the interview sessions, although he provided a rationale for his decisions when questioned about particular examples. As with the preliminary data, the expert used intermediate classifications where he thought appropriate, using up to two gradations from the 'base'

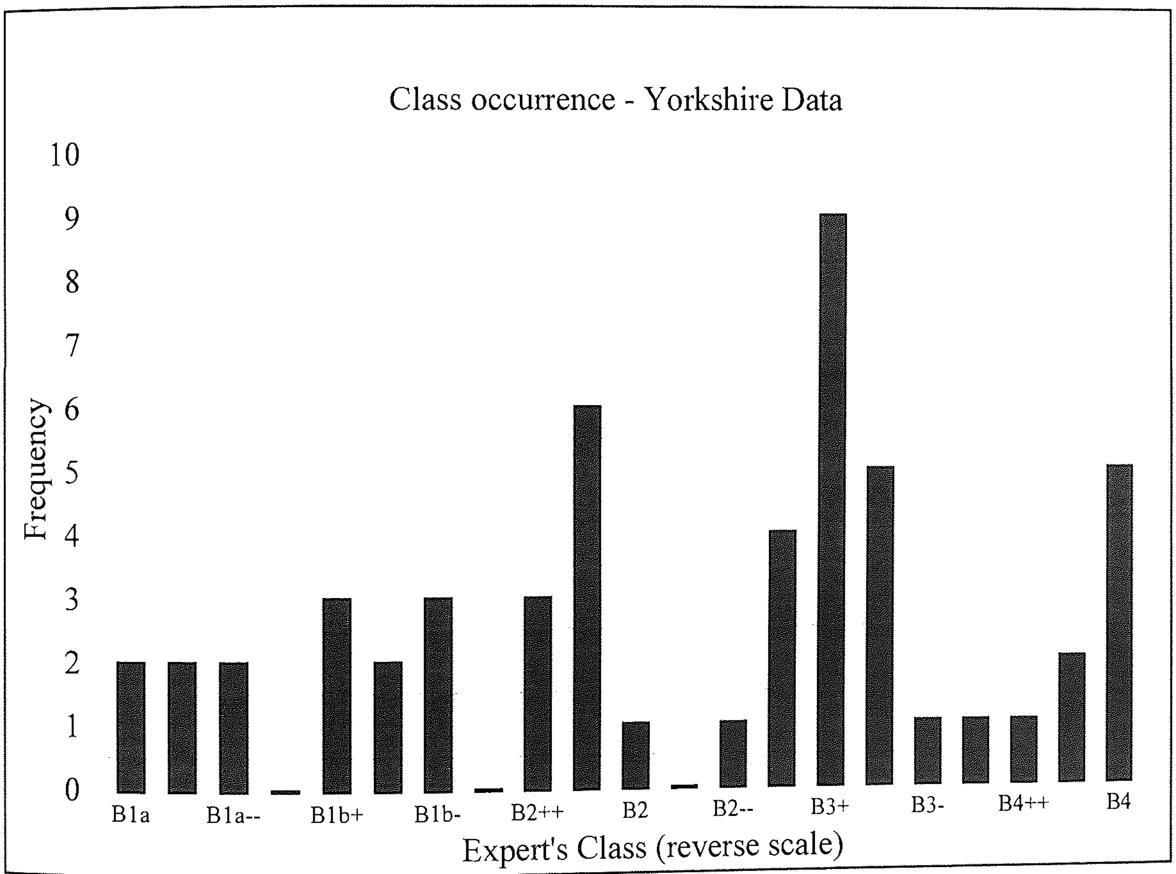

Figure 5.2 Frequency of biological river water quality classes within Yorkshire Data set (intermediate classifications).

classes. Representation of the intermediate classes within the 53 sites is shown in **Figure 5.2**, where the "reverse scale" implies that '+' or '++' denote improving quality. Clearly B3 sites are over-represented in the set, but this is commensurate with the predominant biological quality of the Yorkshire rivers from which the majority of sites are drawn.

5.3.7 Comparison of Expert's and other classifications

The DoE classifications for the Aire and Calder sites provide a means of gauging the correlation between the subjective classifications of the expert and those of YWA biologists (**Figure 5.3**). Caution is required however since for the two classification systems (a) the correspondence between them is unknown and (b) both are ordinal scales, in which the distance between classes is also unknown. For this data set, class A (DoE) as allocated by YWA biologists extends from the expert's B1a to B1b-, with similar ranges for the B, C and D.

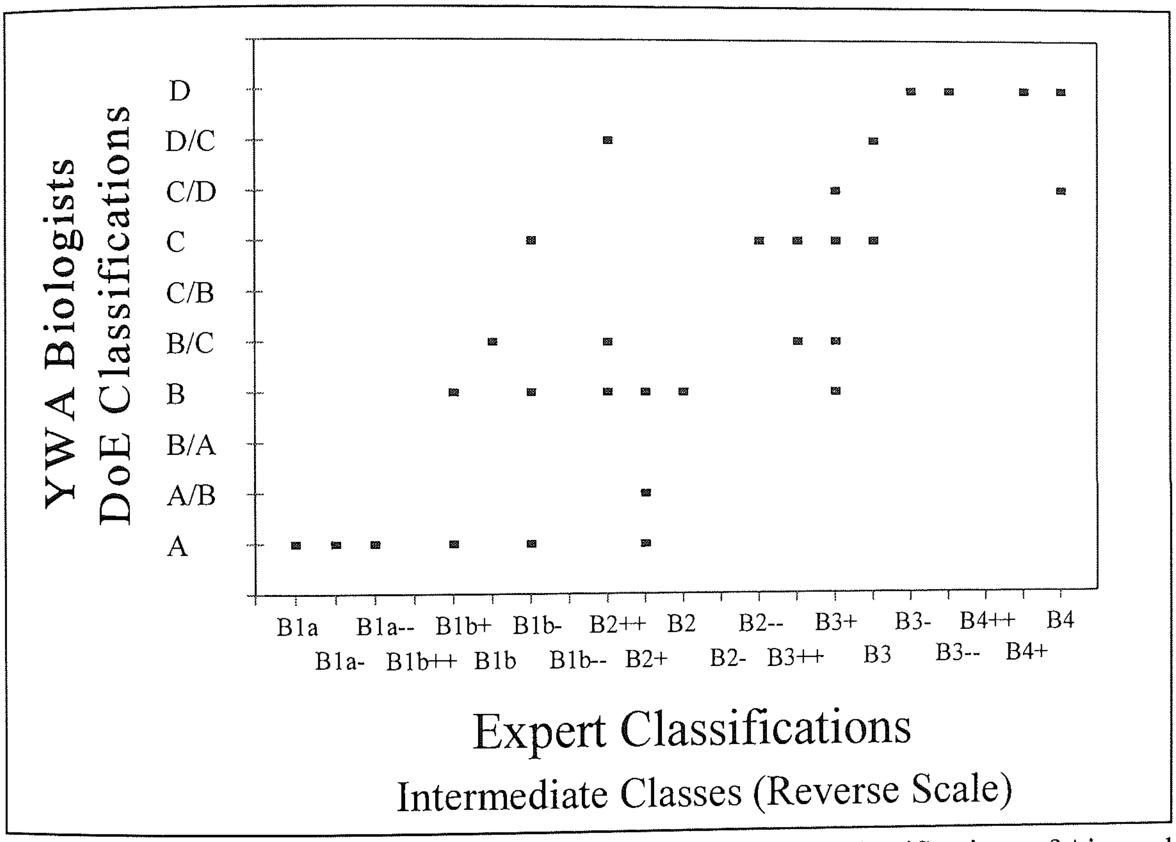

Figure 5.3 Expert's Classifications and YWA Biologists' DoE Classifications of Aire and Calder samples (Note: Distance between classes is unknown).

A more informative comparison is gained from calculating biotic indices for each sample in the entire data set (i.e. all 53 sites). In carrying out this comparison, the expert's classification is taken to be the 'reference' measure: i.e. the standard against which the biotic indices are compared. Mean, minimum and maximum saprobic indices, LQI values, and Chandler and BMWP scores were calculated and plotted against the expert's classifications, expressed in terms of the base biological classes (**Figures 5.4** to **5.7**). The graphs highlight the disparity between discrete and continuous classification schemes. Mean values for the discrete systems LQI and ASPT (not shown) decline uniformly with increasing biological quality, or, for the saprobic index, increase. All the indices show a large overlap in their ranges, when compared against the reference biological classes. Thus, if the expert classification is taken as the reference, it would be difficult to discern changes in water quality from these continuous scores.

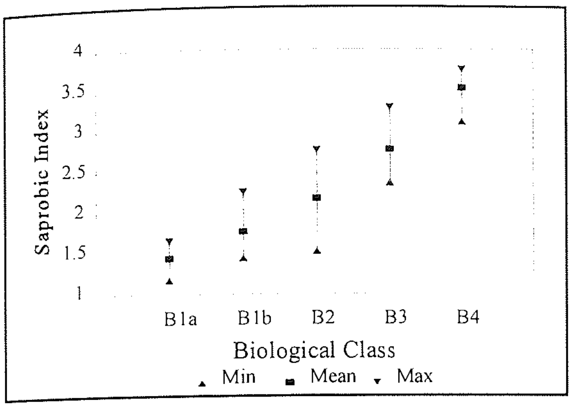

Figure 5.4 Saprobic indices for the biological classes: Yorkshire Data

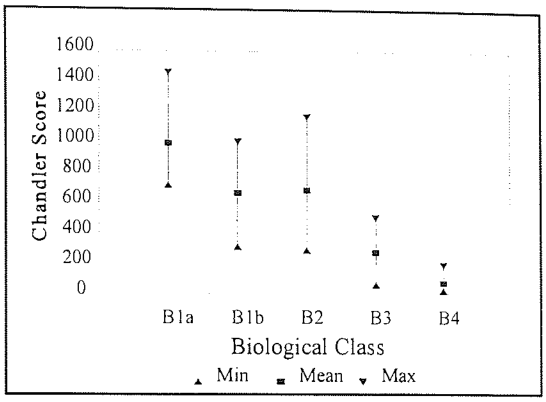

Figure 5.5 Chandler scores for the biological classes: Yorkshire Data

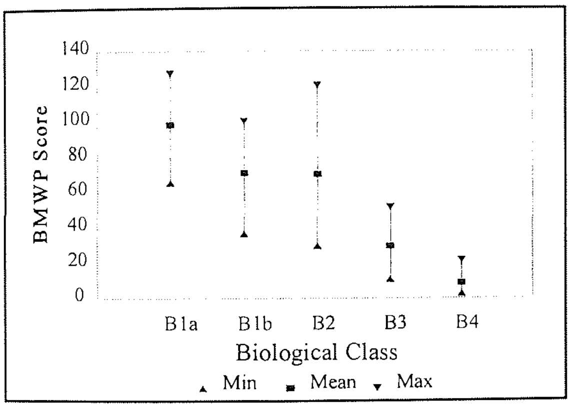

Figure 5.6 BMWP scores for the biological classes: Yorkshire Data

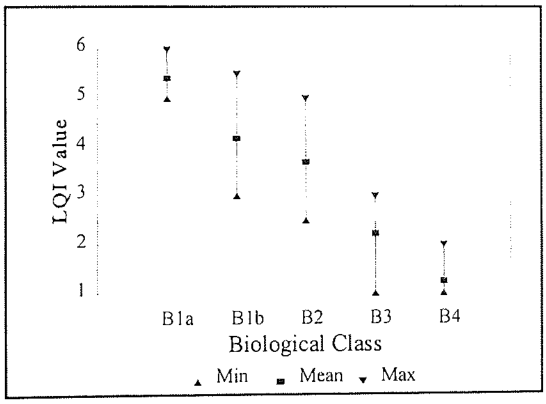

Figure 5.7 LQI values for the biological classes: Yorkshire Data

Several of the samples classified by the expert as B2 were diverse, implying that the diversity of a sample does not always signify high-quality waters (**Figure 5.8**). This is reflected in the large overlap of BMWP scores for B1a, B1b and B2 (**Figure 5.6**). Sites classifed around B1b exhibit a higher diversity than those near to B1a. This is consistent with the expert's viewpoint that the slight increase in organic enrichment as water quality decreases from B1a to B1b can promote productivity. After B2, the general trend is for diversity to decrease.

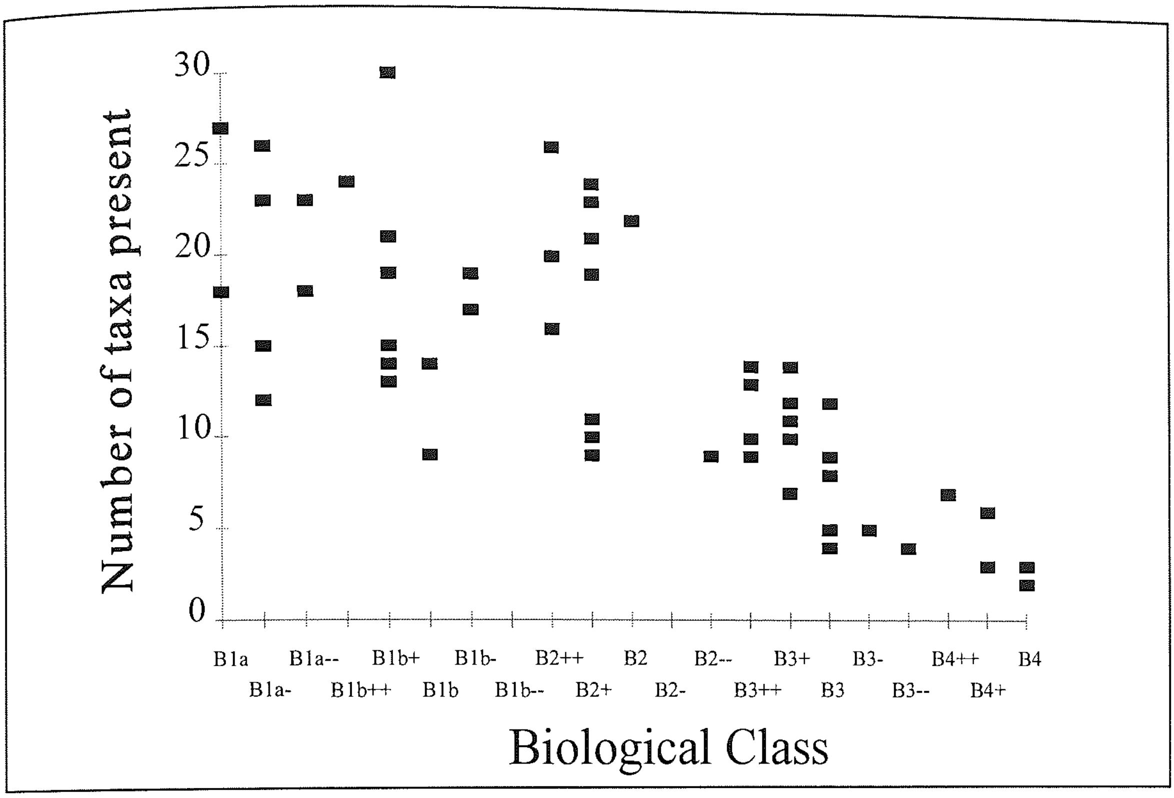

Figure 5.8 Diversity of Yorkshire samples, as measured by number of taxa present, plotted against the expert's classifications

5.4 Severn-Trent Benthic Data

5.4.1 Origin of data

Towards the end of the project, biological surveillance data from the NRA Severn-Trent region was made available to the team of researchers at Aston University investigating the use of AI methods for water quality assessment. The construction of the database of 293 benthic samples from the supplied data, with a detailed description and analysis, was carried out as part of a related project in biological classification using neural networks (Ruck, 1995). Unlike the collated Yorkshire data, the Severn-Trent database contained physical and geographic information on each site, and calculated biotic indices (BMWP, ASPT and TBI scores) for the sample list of invertebrates. The benthic samples were listed at mixed taxonomic levels, with abundance recorded as *present*, *few*, *common*, *abundant* and *very abundant*.

5.4.2 Problems in data utilisation

It was considered at first by the author that this more comprehensive database would be suitable for the series of classification experiments using uncertain reasoning methods. Experiments using subsets of the constructed database were in fact carried out to supplement the results from the Yorkshire data. However a subsequent analysis of how the database was constructed and the samples classified revealed that the NRA data as used in the classification experiments was markedly different from the Yorkshire data in several respects. The forty-one indicator taxa were chosen by the expert from his experience and knowledge and by scrutiny of representative benthic invertebrate samples. Data from the Yorkshire region, supplemented by samples from the Midlands, were used both to construct the Yorkshire data set and expand the group of indicator taxa to forty-one. In this sense, the construction of the Yorkshire data set was an extension of the knowledge elicitation phase of the project.

A study of the Severn-Trent data revealed that the frequency of the forty-one taxa, as chosen by the expert, was lower than expected. To maximise use of the data, the indicator group was modified to admit new taxa or to adjust the taxonomic level, from (say) species to family level. For classification, the expert was shown the full sample list, rather than the abundance levels of the indicator group, as the author originally thought. This misapprehension arose from the fact that the classification exercise for the Severn-Trent data was carried out as part of the project using neural networks, and the format in which the samples were shown to the expert was unknown to the author at the time.

Consequently the output from an automatic classifier whose knowledge is in the form of probability distributions for the original forty-one indicator taxa cannot be directly compared with the expert's classification. As discussed earlier, a decision was made to present both the automatic classifiers and the expert with the states of the indicator group, so that both would be using the same information. For the NRA data, the expert's classification could be influenced by the presence or absence of taxa outside the indicator group. For many sites, representation of indicator group taxa within samples was low, even for B1a and B1b sites. This was not accounted for when running the classification experiments for samples from the Severn-Trent database.

If time permitted, the Severn-Trent data would allow an investigation of the ability of the reasoning systems to classify using only the evidence provided by the states of the indicator taxa, compared with the expert classification of complete, real samples. To make proper use of the Severn-Trent data, further work would be required to account for the representation of the indicator group within each sample, and possibly to increase the size and composition of the indicator group handled by the classifiers. For chronological reasons this work is outside the scope of this project. Therefore, the results from the NRA experiments will not be considered further in this thesis.

5.5 Summary

This chapter has discussed the expert's use of the biological classification system in the context of constructing two databases of benthic samples. The classification of preliminary data, based on just ten indicator taxa, suggested that the expert was content with the practice of biological classification, adapting it to incorporate gradations of river water quality. The adoption of a reference set of forty-one indicator taxa coincided with the development of a more extensive database of benthic samples primarily from the Yorkshire region. The construction of this database has been described, and a descriptive analysis presented.

Biotic indices calculated for each sample were compared with the expert's discrete classifications, which were taken as reference measures of quality. As expected, continuous score systems such as the BMWP and Chandler scores show a high degree of overlap between ranges for the expert's classifications. Although sample diversity showed a general downward trend, B2 sites showed high diversity, suggesting that this measure in itself was not enough to convince the expert of high quality.

Data comprising the 'Yorkshire data set' form the basis of the classification experiments described in the following chapters. The data set suffered from several defects, arising from its storage format (typed reports), incomplete knowledge on sampling methods and sites, inconsistent or mixed modes of taxonomic identification, and from the amalgamation of individual sample data. Consequently the data required filtering into a regular format suitable for classification by both expert and subsequently, automatic classifiers. The methods for filtering this data were described. The information loss arising from this data transformation is contained within the filtering process. To make use of other data sets, a similar filtering process would need to be undertaken before presentation of samples to a domain expert, or the degree of representation of the indicator group within each sample would have to be incorporated into the uncertain reasoning model. This would facilitate an investigation of the ability of the reasoning systems to classify against the expert's classification using complete sample data.

Despite its deficiencies the Yorkshire data set is valuable in that it records the outcome of an important exercise: the biological classification of river water quality based on invertebrate samples by an acknowledged expert in the field. The subsequent chapters will demonstrate that decision algorithms based on methods of uncertain reasoning can emulate the expert's ability to directly classify biological water quality from benthic samples.

Chapter 6

Classification Experiments - I

6.1 Introduction

This chapter describes the first part of the work undertaken to develop "automatic classifiers" for river water quality classification. These are computer programs incorporating decision algorithms drawn from methods of uncertain reasoning which attempt to emulate the expert's ability to determine the quality of river water from samples of benthic invertebrate communities. In this chapter the performance of classifiers in which belief is represented as Bayesian or singleton support is examined. The effect of varying evidential weight is investigated in a series of computational experiments, both for emulating the expert's own weighting of evidence and of managing evidential conflict.

The performance of the classifiers under varying decision and conflict threshold levels is also tested. The chapter begins with a discussion of the environment within which the classification computer programs were developed. It then describes the preliminary systems that formed a pilot study into the use of uncertain reasoning algorithms for biological classification, and discusses the mechanisms adopted for assessing the performance of the classifiers.

6.2 Background to experiments

6.2.1 Development environment

Computer programs for automatic classification were developed on a personal computer initially in both FORTRAN and the LEONARDO expert system shell. This latter system provides a convenient environment for developing the automatic classifiers, with its facilities for constructing user-interfaces and the ability to integrate algorithmic, causal and object knowledge. It was therefore adopted as the primary development platform for most of the classification experiments. It was also envisaged that the use of this development environment would allow further development of the BERT expert system.

Knowledge within a LEONARDO application can be represented in several ways: in the familiar production rule format, and in the case of object knowledge as frames. Every LEONARDO knowledge base consists of a group of production rules known as the Main RuleSet which is compiled to produce a list of objects used by those rules. These objects can either have values of a certain type (textual, numerical or a list of textual information) or be associated with procedural or user-interface code. Normally a text object can only have one value, unless it is the goal of a knowledge base reasoning under uncertainty and representing multiple hypotheses. In this case the goal object is multi-valued, each value having an associated certainty in the range 0.0 to 1.0. Procedural knowledge can also be encoded in much the same way as a conventional subroutine or function. The development environment allows knowledge bases to be developed rapidly, and the system provides facilities for effective and convenient user-interface design. The default user-interface can be extended to include screens and help information written as hypertext customised to particular requirements.

6.2.2 Early system development

6.2.2.1 Overview

In the early stages of the project the processes of knowledge elicitation and system development were interactive and essentially concurrent. Simple classification systems were produced quickly to increase familiarity with Bayesian analysis and to investigate the efficacy of the elicitation exercise itself, which was still largely exploratory, and the feasibility of using uncertain reasoning methods for biological classification. This section describes this preliminary work.

Results from the early classification systems were fed back to the domain expert in the knowledge acquisition sessions, whose comments helped improve the method of eliciting the probability distributions and the quality of the data obtained. This cycle was repeated several times before system development could continue independently of the knowledge acquisition sessions. Continual liaison with the domain expert was however necessary to test the performance of the classification systems.

The first decision algorithm considered was the simplified Bayesian method in its odds-likelihood form. Initially, the requisite likelihoods of sufficiency and necessity referred to the states of presence and absence for each of the original ten indicator taxa described in Chapter 4. Among the reasons for choosing the Bayesian odds-likelihood formulation was the historical prominence of this calculus (as described in Chapter 3) and the facility for Bayesian reasoning in LEONARDO.

6.2.2.2 Results for two abundance states

Classifiers executing under LEONARDO's Bayesian control present multiple values of the goal object (in this case water quality) in a rank order of certainty for each of the corresponding five hypotheses, for example: B3(0.98), B4(0.88), B2(0.15), B1b(0.00), B1a(0.00). The results for the five preliminary data sets are shown below in **Table 6.1**. Each item of evidence from the first ten indicator taxa is considered to exist in just two states: present or absent. As explained in Chapter 5 the expert was asked to grade the samples with reference to the biological classes B1a to B4. In three cases the domain expert, acting on his own initiative, used intermediate classifications. Rather than use the rank ordering produced by LEONARDO, the probability values are shown in order of the classification system from B1a to B4. Note that the values are unnormalised.

Table 6.1 Summary of Preliminary Bayesian classifications for two states

Data ref.	**Classifier output**					**Expert Class.**
	B1a	**B1b**	**B2**	**B3**	**B4**	
Prelim 1	0.00	0.00	0.04	0.93	0.97	B3--
Prelim 2	0.00	0.00	0.02	0.94	0.97	B3
Prelim 3	0.98	0.96	0.02	0.00	0.00	B1b
Prelim 4	0.00	0.00	0.15	0.98	0.88	B3-
Prelim 5	0.00	0.00	0.94	0.92	0.00	B3+

Note: System classifications are unnormalised. Expert classifications refer to 'reverse' scale, in which (-) denotes decreasing river water quality, (+) increasing quality. See text for explanation.

In **Table 6.1** data sets Prelim 3 to Prelim 5 relate to actual results obtained from a biological assessment of the effect of an organic load being discharged into a river. Prelim 3 refers to a sample from a station above the discharge, Prelim 4 to a station immediately downstream, and Prelim 5 to one further downstream. In these three cases the classifications obtained from the LEONARDO application were consistent with the expert's assessment.

6.2.2.3 Improved Classifiers using Abundance-levels

At this stage, these results were used primarily to show the feasibility of Bayesian classification from biological data. The conclusion from this study was that the method was feasible and worthy of further development. However as explained in Chapter 4 the manner in which values of L_s and L_n were encoded from verbal expressions of likelihood was unsatisfactory. By means of graphical pro-forma, the domain expert could indicate the likelihoods of the biological classes given the sensor evidence directly. This evidence was itself refined from two states of presence and absence to four: rare, established, abundant and absent.

The Bayesian knowledge bases were modified to incorporate an extended rule set using these new values for the four sensor states. **Table 6.2** shows a summary of results for the preliminary data using an improved classification system that contains the full rule-set.

Table 6.2 Summary of Preliminary Bayesian classifications using abundance-levels

Data ref.	**Classifier output**					**Expert Class**
	B1a	**B1b**	**B2**	**B3**	**B4**	
Prelim 1	0.00	0.00	0.00	0.94	0.97	B3--
Prelim 2	0.00	0.00	0.02	0.99	0.00	B3
Prelim 3	0.91	0.97	0.01	0.00	0.00	B1b
Prelim 4	0.00	0.00	0.00	0.91	0.96	B3-
Prelim 5	0.00	0.00	0.81	0.99	0.00	B3+

6.2.2.4 Effect of adjusting probability distribution

The preliminary experiments also allowed for two separate probability distributions. As described in Chapter 4 the elicited probability measures for the indicator taxa were adjusted to form new histograms with redistributed probability mass. The bar heights of these adjusted histograms correspond to the areas subtended by a line joining the midpoints of the original bars. The net effect is to reduce the crispness of the derived discrete probability values. The differences in support for each hypothesis were noticeable, and suggested that the effect of this adjustment should be investigated.

6.2.2.5 Conclusions from early classifications

The small size of the preliminary data-set precludes any definite conclusions to be drawn from the pilot study, except to show the feasibility of this approach. However, the results did generate questions deemed worthy of further investigation:

(1) how important was the nature and strength of the evidence as represented by the probability distributions?
(2) how should the performance of the classifier be assessed?
(3) how should decisions be made regarding the overall classification, i.e. how should one interpret the results of the classification?
(4) which decision algorithms should be used to combine evidence?

Of particular importance is the interpretation of the output from the classifiers. This is considered in the next section.

6.3 Decision mechanisms

6.3.1 Intermediate classifications

In interpreting the output from an automatic classifier one has to consider exactly what is being measured, and with what its performance is being compared. To do this, the expert's assessment of river water quality using the biological classes is briefly reexamined. The biological classification system adopted by the domain expert and the automatic classifiers is essentially an ordinal or ranked scale of quality measurement in which the intervals between the classes are unknown. However class B1a is considered "better quality" (or less organically enriched) than B1b, and so on.

The expert was asked to classify river water quality in terms the biological classes B1a to B4. From the outset however, the expert used intermediate classifications, initially using '+' or '++' to show degradations in quality from that implied by class boundaries, and '-' or '--' showing improvements in quality, and the class-boundary values.[1] Thus 'B2+' denoted a classification by the expert of B2 veering towards B3, i.e. poorer quality than B2 itself. This procedure derives partly from the indirect correspondence between this classification and the saprobic systems in which high-quality waters (e.g. xenosaprobic or oligosaprobic) receive low numerical values for saprobic indices, and poorer quality waters (such as mesosaprobic or polysaprobic waters) receive higher values. As employed by both Sládeček (1973) and Pantle and Buck (1955), the saprobic indices derived from taxa lists were mapped back to the saprobic zones by the adoption of a linear interval scale

---

[1] This practice of using intermediate classifications was also employed by the Yorkshire Water Authority and is continued by its successors, the NRA (Northumbria and Yorkshire Region) (now the Environment Agency).

corresponding to increasing organic enrichment. Indices in the range 1.0 to 1.5 mapped to the oligosaprobic zone, the range 1.5 to 2.5 to ß-mesosaprobic, and so on.

Given this practice and the fact that the biological classification system is a measure of organic pollution it is tempting to map the expert's classifications to a linear scale to compare the intermediate classifications with the automatic classifier output. Such a linear mapping was in fact carried out in an earlier analysis. The 'base' classes B1a to B4 are weighted by a factor corresponding to increasing organic enrichment, analogously to the saprobic scale as used by Sládeček, so that the weights for each class are 0,1,2,3,4 respectively. The expert's intermediate classifications were mapped by considering two increments or decrements (denoted by '++' and '--' respectively) to correspond to river water quality approximately midway between two classes.

Each incremental indication corresponds to an increment or decrement of 0.2 on this scale. A classification by the expert of B2- would therefore correspond to an "Expert Classification Index" (ECI) of 1.8, B3++ to an ECI of 3.4 and so on. In previous work, this measure was compared directly with the "System Classification Index" or SCI (explained below) to calculate the mean square error for a classification over T samples:

$$MSE = \sum_{s=1}^{T} |(ECI - SCI)_s|^2 \;/\; T \qquad \textbf{(6.1)}$$

where $(ECI - SCI)_s$ is the difference between the expert's and system's classifications for site s. This measure was then used to show the deviation of the automatic classifier's decision from that of the expert. One objection to this approach however is that it assumes the existence of an interval scale for biological quality, violating the adoption of discrete, ordinal categories. The *MSE* or *ECI* cannot be used if there does not exist a linear "biological quality scale" underlying the classes. The resolution to this problem is considered in the next section.

6.3.2 System Classification Index

The Bayesian and Dempster-Shafer theories assign degrees of support, as probability mass, to the competing propositions within the sample space. Strictly, since these are mutually exclusive, only one can be true. In practice however we can never know which one is definitely true unless all uncertainty has been removed and support for all but one proposition reduces to zero. Since this rarely obtains, we need to deal with situations in which rival propositions have non-zero degrees of support after all the evidence has been combined.

Consider a situation in which the output from a Bayesian classifier suggests that, after combining all the available sensor evidence, the support for the classes B1a, B1b,B2,B3,B4 is (0.0,0.75,0.25,0,0). In this example quality class B1b receives majority support of 0.75, far exceeding its nearest competitor B2 that receives 0.25. All other classes receive zero support. The output is essentially a rank-ordered set of probabilities, normalised to sum to unity, presented as support for each of the quality classes.

On examination of the output, one may adopt a "highest-wins" decision strategy and consider that the system's classification is simply B1b, dismissing all other alternatives. While support for B2 is much smaller than B1b however, it is not negligible. Adjacent hypotheses may receive comparable support, making a "highest-wins" decision mechanism difficult (in deciding which hypothesis should win) or undesirable (in dismissing plausible alternatives).

Maintaining parallel chains of reasoning for as long as possible is more important than deciding when to dismiss one or more of the hypotheses. The ability to support multiple plausible chains of reasoning is a primary reason for using uncertain-reasoning methods. Ideally therefore the support for each of the competing hypotheses should participate in the overall decision by the automatic classifier. Therefore a system classification index that took account of the support across the competing hypotheses was designed, which attempted to reflect the expert's own tendency to recognise intermediate classifications.

For summary purposes only the multihypothesis support from the classifiers can be expressed via the *system classification index* (SCI) defined as:

$$SCI = \sum_{A \in \Theta} w_A . S_A \tag{6.2}$$

where w_A is the numerical weight for class A in the frame of discernment Θ, and S_A is the computed support for that class. The SCI is therefore similar to the calculation of the saprobic index, with the difference that the later calculates a weighted average of the support for each saprobic zone. For support of (0.0,0.75,0.25,0,0) the system classification index is from **(6.2)**: 0 x 0.0 + 1 x 0.75 + 2 x 0.25 + 3 x 0.0 + 4 x 0.0 = 1.25, using the numerical weights given in section 6.3.1. This figure can be interpreted as a classification close to B1b, between this class and B2.

The *SCI* index works well for the majority of classifier output in which the distribution of probability mass is unimodal, which shows a high degree (>3) of kurtosis, with the skewness varying with the biological class receiving dominant support. Problems arise for certain rarely-occurring output distributions - those with zero information content, for example (0.2,0.2,0.2,0.2,0.2) or for bimodal distributions that result in a noncontiguous rank-ordering of the classes. For example, a distribution (0.4,0.0,0.0,0.2,0.4) results in an *SCI* of 2.2, incorrectly assigning this output between classes B2 and B3. These contradictory or vacuous distributions do not occur often in classifier output, but a simple index was developed to indicate how much confidence one can have in a classifier's decision.

6.3.3 Indicator value

An indicator value *I* was calculated for each classifier output that accounted for the information content of the distributions and the rank ordering of class support. For a decision maker, the information content should be high (i.e. it should be clear which of the classes have majority support), and the first two class ranks in a rank ordering of support should be contiguous. As an example, the distribution (0.2,0.7,0.1,0,0) has both a high information content (due to the large degree of support for class B1b) and contiguously orders the first two ranks: (B1b,B1a,B2,B3,B4). In contrast the distribution (0.4, 0.0,0.0,0.2,0.4) has low content and a rank-ordering (B1a,B4,B3,B1b,B2) in which the noncontiguous class-support for both B1a and B4 simultaneously is contradictory.

The *I*-value used for the classifier output distributions was similar to that used for the indicator taxa, except that a normalisation factor was introduced so that the vacuous distribution (0.2,0.2,0.2,0.2,0.2) has an *I*-value of unity. This represents the minimum information state, corresponding to the *a priori* probabilities assigned to the hypotheses before any evidence is considered. With the *SCI* and the decision order, the *I*-value should allow a decision-maker to accept or reject the output from a classifier using an uncertain-reasoning algorithm. Low *I*-values (say < 10) could be rejected as unsafe classifications.

6.3.4 Intermediate classification mappings

In an attempt to interpret the *SCI* in terms of the expert's intermediate classes, it was postulated that the latter could be viewed as having partial membership of the 'base' biological classes B1a,B1b,B2,B3,B4. This approach is analogous to determining membership of fuzzy sets (Zadeh, 1965). This contrasts with the original mapping of the expert's intermediate classifications to a linear scale which, it was previously suggested, is unjustifiable for an ordinal classification system. The expert's use of B1b++ or B2- is problematical in that while these assessments are between the base classes there is no indication of their relative position. Improving and degrading quality increments (e.g. between B1b-- and B2++) may also overlap.

Table 6.3 shows one of many possible membership assignments to the range of intermediate and base classes. These membership grades are, as with Bayesian methods, a matter of subjective belief. However, the assignments to (for example) B1a-- and B1b++ are chosen to reflect the fact that distinguishing between a poorer-quality B1a and an exceptional B1b river site would be difficult. The *SCI* for single intermediate divisions (e.g. B1a- or B2+ and so on) resulting from these mappings corresponds to those used in an associated classification study of NRA Severn-Trent benthic data using artificial neural networks. The membership grades for each of the base classes are of course unity with respect to the corresponding class. Using this table, assigning an intermediate class to a classifier's output was then possible based on its *SCI*. This measure was used only for individual site classifications: the performance over the entire data-set was assessed via classification rates, explained below.

Table 6.3 Suggested membership grades for intermediate classifications of the base classes. The base classes are {B1a,B1b,B2,B3,B4}

Intermediate Class	Base Biological Class					
	B1a	B1b	B2	B3	B4	SCI
B1a	1	0	0	0	0	0
B1a-	0.67	0.33	0	0	0	0.33
B1a--	0.55	0.45	0	0	0	0.45
B1b++	0.45	0.55	0	0	0	0.55
B1b+	0.33	0.67	0	0	0	0.67
B1b	0	1	0	0	0	1
B1b-	0	0.67	0.33	0	0	1.33
B1b--	0	0.55	0.45	0	0	1.45
B2++	0	0.45	0.55	0	0	1.55
B2+	0	0.33	0.67	0	0	1.67
B2	0	0	1	0	0	2
B2-	0	0	0.67	0.33	0	2.33
B2--	0	0	0.55	0.45	0	2.45
B3++	0	0	0.45	0.55	0	2.55
B3+	0	0	0.33	0.67	0	2.67
B3	0	0	0	1	0	3
B3-	0	0	0	0.67	0.33	3.33
B3--	0	0	0	0.55	0.45	3.45
B4++	0	0	0	0.45	0.55	3.55
B4+	0	0	0	0.33	0.67	3.67
B4	0	0	0	0	1	4

6.3.5 Classification Rates

For an individual decision problem several indices were used to assess the performance of a site classification: *SCI*, *I*-value, nearest intermediate and whole classification, and the probability masses and decision order for the five hypotheses. Using these indices, a decision maker can decide, from inspection, whether the classification's decision was acceptable or not. To assess the performance of a classifier over the data-set, a criterion was adopted for deciding whether a classification error had occurred. The *empirical error rate* is defined as the number of errors divided by the number of sample cases and the *classification rate* as (1 - *Error-rate*). A successful classification was deemed to occur if, for a particular benthic sample, the expert's opinion and the classifier's decision both align on the same base class. This criterion allows the use of the *confusion matrix* in assessing performance over repeated classifications. The confusion matrix is identical to the contingency table used in non-parametric statistics in which the degree of dependence between the rows and columns of the table can be determined. This can be seen directly from inspection of the table, or from the classification rate. Alternatively non-parametric measures such as Spearman's rank coefficient can be used to quantify correlation between the rows and columns of the table.

For a n-class problem, the confusion matrix or contingency table is of dimension $n \times n$. The number of classification errors is readily determined by summing the off-diagonal elements in the matrix. A perfect classifier would result in a diagonal matrix, with zero error rate. The confusion matrix can be used directly in decision making if certain empirical measures of utility, risk, or cost exist. For instance, one could consider the <u>cost</u> of misclassification. For a two-class Positive/Negative problem, the confusion matrix has four elements:

$$\begin{pmatrix} TP & FP \\ FN & TN \end{pmatrix} \tag{6.3}$$

where FN refers to a "False Negative" decision and FP to "False Positive".

For the biological classification problem, the "false negatives" correspond to the elements to the left and below the diagonal of the confusion matrix. These represent those cases in which the automatic classifier, based on the sample data, computed the water quality as poorer than that decided by the expert, i.e. the system erred on the poorer side. The "false positives" are those above and to the right of the diagonal, corresponding to system classifications of higher quality than the true class. For most practical problems, the costs associated with false positives and false negatives will be different. Although this information for the biological classification problem does not currently exist, one can envisage how the costs of misclassification may be apportioned. In considering the design of the BMWP score for instance, the most tolerant taxon within each scoring group is selected for awarding points towards the overall score, biasing against over-optimistic quality assessments. This bias would imply that erring on the side of poorer quality is "safer" than the converse, i.e. the "costs" of misclassifying river water quality higher than it is being greater than those due to incorrectly downgrading quality.

Thus one could construct a 5 x 5 misclassification cost matrix M reflecting the higher costs above the diagonal than below, and the fact that costs usually increase with misclassification. The diagonal elements M_{ii} are all zero since there is no cost-penalty for correct classification. A (fictitious) cost matrix could be as shown in shown in **Table 6.4**.

Table 6.4 Form of possible cost matrix for biological classification problem

$$\begin{pmatrix} 0 & 3 & 4 & 5 & 6 \\ 2 & 0 & 3 & 4 & 5 \\ 3 & 2 & 0 & 3 & 4 \\ 4 & 3 & 2 & 0 & 3 \\ 5 & 4 & 3 & 2 & 0 \end{pmatrix}$$

The actual misclassification costs are then computed by:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} M_{ij} \qquad \textbf{(6.4)}$$

where C is the n x n confusion matrix. The figures in **Table 6.4** are arbitrary, and in a practical application of biological classification they would have to be determined empirically. This method illustrates one other way in which the performance of a classifier could be assessed.

6.4 Experiments using Bayesian and Singleton-support functions

6.4.1 Representation of Bayesian belief

Classes of belief functions effectively constrain the possible mass assignments in numerical reasoning schemes. As an example, consider the belief function representing total ignorance. The so-called *vacuous belief function* $Bel(A) = 0$ $(A \neq \Theta)$ implies that no information is available regarding the truth of propositions in the power set, apart from the frame of discernment itself for which $Bel(\Theta) = 1$. In Bayesian theory ignorance is expressed using the principle of indifference by assigning an equal distribution of mass across the singleton hypotheses, although this action is not necessarily devoid of information. The Dempster-Shafer representation of ignorance is more satisfactory in that none of the subsets of Θ receive any credence. In contrast, the *certain belief function* $Bel(A_i) = 1$ implies that the proposition represented by the subset A_i is true beyond doubt, so that $m(A_i) = 1.0$ and $m(A_j) = 0$ for all $j \neq i$. If A_i is a singleton, the evidence is completely decisive.

Bayesian belief functions are equivalent to a conventional discrete probability distribution across the singleton hypotheses H_i. Therefore

$$\sum_{H_i \in \Theta} m(\{H_i\}) = 1.0 \tag{6.5}$$

Also since $Bel(A_i) = m(A_i)$ for singleton hypotheses, and the width of the belief interval is zero, we have

$$m(\{H_i\}) = Bel(\{H_i\}) = Pls(\{H_i\}), \tag{6.6}$$

for all $H_i \in \Theta$. For Bayesian belief functions, Dempster's rule is equivalent to the Bayesian updating mechanism. Moreover, if a Bayesian belief function is combined with a non-Bayesian belief function, the resulting function either is Bayesian or does not exist (Voorbraak, 1990). So any uncertainty in the system that exists beforehand is eliminated when updated by a Bayesian belief function.

6.4.2 Implementation of Bayesian belief

The assignment of probability mass to the competing hypotheses within the frame of discernment is dependent on the reasoning algorithm used and the representation of belief. For Bayesian reasoning, a direct correspondence exists between the probability distributions and the support for each of the singleton hypotheses, the river water quality classes. Two variants of the probability distributions were used: one derived directly from the histograms elicited from the domain expert, and the second obtained by applying a smoothing adjustment to the distributions. Within the classifiers, an option was provided for the user to select the form of the distribution. One can use either $P(H|e)$ or $P(e|H)$ to assign Bayesian belief, since the normalising constant in Bayesian updating can be directly computed by requiring the $P(H_i|e)$ to sum to unity over H_i.

With Bayesian reasoning the representation of belief is essentially that of singleton support: this therefore largely predetermines the assignment of probability mass. As with Dempster-Shafer reasoning a degree of evidential discount can be applied to the sensor evidence. However with Dempster-Shafer reasoning, applying evidential discount ϵ to singleton support results in uncommitted support residing with the environment, and likewise support for each singleton is reduced by a factor $(1 - \epsilon)$. For Bayesian reasoning this facility of leaving belief uncommitted does not exist: instead, the effective probability mass m'_i for a singleton hypothesis H_i given evidential discount ϵ becomes:

$$m'_i = m_i (1 - \epsilon) + \epsilon / n \tag{6.7}$$

for n hypotheses. The effect is to redistribute the probability mass over the singleton sets similar to the distributions obtained from the area-adjustments. Like this latter adjustment, the effect of evidential discounts in Bayesian reasoning is to produce a less crisp representation of the sensor evidence. For $\epsilon = 0$, the support for each hypothesis corresponds to the probability distribution read from the spreadsheet data; if $\epsilon = 1$, the sensor evidence contributes no information ($m'_i = \epsilon/n \ \forall\ i$), i.e. its evidence is completely discounted.

6.4.3 Representation of Singleton support

Singleton support is a term used by this author to denote an assignment scheme in which the frame receives a proportion of probability mass equal to the degree of evidential discount. This is the amount by which support for each proper subset of Θ is reduced, the remaining mass distributed among the singletons. The idea of discount was described in Chapter 3. It was proposed in Shafer's original treatise for dealing with evidential conflict (Shafer, 1976). By introducing a discount rate for each proper subset, the "crispness" or certainty of the evidence for the equivalent propositions is reduced, and the resulting uncertainty in the system represented by a mass assignment to Θ. Consider the Bayesian belief mass assignment of the form

$$\begin{gathered} m(\{\mathrm{B1a}\}) = a,\ m(\{\mathrm{B1b}\}) = b,\ m(\{\mathrm{B2}\}) = c, \\ m(\{\mathrm{B3}\}) = d,\ m(\{\mathrm{B4}\}) = e \end{gathered} \tag{6.8}$$

If now $m(\Theta) = \epsilon$, then each of the basic probability numbers is reduced in proportion, e.g.

$$m(\{\mathrm{B1a}\}) = a(1-\epsilon),\quad m(\{\mathrm{B1b}\}) = b(1-\epsilon)\ldots \tag{6.9}$$

Note that if the discount rate $\epsilon = 1$, the belief function becomes vacuous, with no information content whereas if $\epsilon = 0$, the mass assignment for the singleton support belief function reduces to a Bayesian belief function.

6.4.4 Implementation of Singleton-support functions

As discussed in Chapter 3 the Dempster-Shafer calculus is a generalisation of the Bayesian approach. If in Dempster-Shafer reasoning a Bayesian belief function[2] is adopted for evidence representation, the scheme is equivalent to Bayesian reasoning. Within Dempster-Shafer reasoning the combination of a Bayesian belief function with a non-Bayesian function either does not exist or results in a Bayesian function. Therefore the representation of any one set of evidence as Bayesian belief during the Dempster-Shafer combination procedure reduces the algorithm to the Bayesian approach even if the remainder of the evidence is represented as singleton support. This would suggest that one could use the Dempster-Shafer calculus for both Bayesian and Dempster-Shafer belief representations. However since evidential discount is treated differently in the two calculi maintaining two different types of classifiers for singleton support was necessary, one for Bayesian, the other for Dempster-Shafer.

6.4.5 Effect of probability distribution adjustment

6.4.5.1 Motivation

These experiments were carried out to test the effect of representing the derived probability distributions for the sensor evidence in each of two modes: one mode directly based on the original graphical pro-forma elicited from the domain expert, the other corresponding to the area-adjusted distributions developed as described in Chapter 4. Adjustments to the probability distributions are equivalent to varying evidential strength by a set amount. The intent of this investigation was to determine whether this adjustment affected classification performance.

6.4.5.2 Procedure

Sites from the Yorkshire Water data were used to test the performance of a Bayesian classifier that could use both modes of the sensor evidence. Rare evidence was discounted from the classification, on the assumption that it was of insufficient 'strength' to influence the decision. The remaining evidence, corresponding to abundant, established and absent taxa and presented to the Bayesian updating algorithm in that order (referred to as "sensor-state" order) was considered at full strength.

---

[2] A Bayesian belief function can be viewed as singleton support with zero evidential discount.

6.4.5.3 Results for data-set

Table 6.5 shows the classification performance over the Yorkshire data using the probability distributions derived directly from the expert, while **Table 6.6** show the classification obtained by using the adjusted distributions.

Table 6.5 Bayesian classification of Yorkshire data using original probability distributions.

Classifier Output		Expert Decision				
		B1a	B1b	B2	B3	B4
	B1a	1	0	0	0	0
	B1b	5	6	0	0	0
	B2	0	2	9	2	0
	B3	0	0	1	18	2
	B4	0	0	0	1	6

Notes: Rare evidence neutral. Classification rate 75.47%.

Table 6.6 Bayesian classification of Yorkshire data using adjusted probability distributions.

Classifier Output		Expert Decision				
		B1a	B1b	B2	B3	B4
	B1a	4	2	0	0	0
	B1b	2	4	1	0	0
	B2	0	2	8	1	0
	B3	0	0	1	17	0
	B4	0	0	0	3	8

Note: Rare evidence neutral. Classification rate 77.36%.

6.4.5.4 Discussion

Using both variants of probability distribution mis-classifications are no greater than one class either side of the expert's base class. However the classification over the data set exhibits a better performance when the distributions are smoothed via area-adjustment. The difference between the classifications when expressed as ranks is significant at the 5% level. Classification for good quality (B1a) samples using the original distributions results in a tendency to downgrade by one class. Overall, the tendency is for the system to err on the poorer-quality side, which is preferable to overestimating the quality of the sampled site.

The effect of adjusting the probability distributions representing the sensor seems to support the observations made during the preliminary system development that the crispness of the original distributions may produce premature decisions. The number of indicator taxa participating in the decision was recorded for each sample. Once a particular hypothesis receives 100% support, all remaining evidence within the indicator taxa group is immediately eliminated. Such decisions are more likely to be reached for highly focused evidence. The number of participating indicator taxa was less than or equal to 10 for 27 out of the 53 samples using the original distributions. For the area-adjusted distributions, all of the forty-one taxa within the indicator group (apart from those whose status was Rare) took part in the classification decision for each sample.

Table 6.7 Evidence combination of indicator taxa for Yorkshire Water site ref. 611 (Original distributions).

Taxon	Sensor-state	B1a	B1b	B2	B3	B4
Potamopyrgus jenkinsi	ESTABLISHED	0.35	0.14	0.30	0.21	0.00
LUMBRICULIDAE	ESTABLISHED	0.00	0.37	0.28	0.35	0.00
Combined Support		0.00	0.25	0.40	0.35	0.00
Baetis rhodani	ESTABLISHED	0.36	0.18	0.21	0.25	0.00
Combined Support		0.00	0.22	0.39	0.39	0.00
Rhithrogena spp.	ESTABLISHED	0.56	0.44	0.00	0.00	0.00
Combined Support		0.00	1.00	0.00	0.00	0.00

Notes: The Bayesian algorithm used the original distributions elicited from the expert.

Table 6.7 shows the evidence considered for a particular Yorkshire site (611), classified by the expert as excellent biological quality (B1a). The first indicator considered is the snail *Potamopyrgus jenkinsi* in the Established abundance state, which suggests highest support for class B1a. After combination with the evidence provided by Lumbriculidae, also Established at the site, majority support now shifts to B2. Note however that support for B1a is eliminated: the evidence from Lumbriculidae effectively vetoing any further support for that class. Although *Baetis rhodani* when Established is a strong indicator of very good quality, this hypothesis has been eliminated by previous evidence. The same is true of the mayfly *Rhithrogena spp.*, whose support is highly focused on classes B1a and B1b when in the Established state, with support for lower-quality classes being zero. The result of combining the previous distribution with this new evidence is that a decisive outcome for class B1b. No further evidence can be considered, so the decision is made with four taxa.

Consider the classification of the same site when the Bayesian classifier is presented with the adjusted distributions provided by the same set of evidence (**Table 6.8**). The smoothing of the original distributions means that support for B1a from Lumbriculidae is non-zero, although very small. However combination of this evidence with that of *Potamopyrgus jenkinsi* results in the maintenance of belief in this class, so that the probability profile after considering sets of evidence is markedly different from that shown in the last row of **Table 6.8**. The Bayesian classifier combines in all the 30 sets of evidence from the indicator group for this site.

Table 6.8 Evidence combination of indicator taxa for Yorkshire Water site ref. 611 (Adjusted distributions).

Taxon	Sensor-state	B1a	B1b	B2	B3	B4
Potamopyrgus jenkinsi	ESTABLISHED	0.32	0.15	0.28	0.22	0.03
LUMBRICULIDAE	ESTABLISHED	0.05	0.32	0.29	0.30	0.04
Combined Support		0.02	0.05	0.08	0.07	0.00
Baetis rhodani	ESTABLISHED	0.36	0.19	0.20	0.22	0.03
Combined Support		0.12	0.19	0.36	0.32	0.00
Rhithrogena spp.	ESTABLISHED	0.54	0.40	0.06	0.00	0.00
Combined Support		0.39	0.48	0.12	0.00	0.00

6.4.5.5 Summary

It appears from the results that varying evidential strength via this adjustment of probability mass does reduce misclassification error for the Bayesian classifier. For Bayesian belief, the probability mass is removed from singletons with higher support and redistributed to singletons with lower support. If the support for a singleton as suggested by the original distributions is zero, adjusting the distribution will lead to it having non-zero support.

Maintaining non-zero support for propositions throughout the evidential combination maintains plausible belief for as long as possible until all the evidence from the indicators have been combined. There is therefore less likelihood of premature decisions with the adjusted distributions that with the more highly-focused evidence.

6.4.6 Variations in Evidential Strength

6.4.6.1 Motivation

In Chapter 3 it was suggested that the process of directly interpreting the benthic samples from the river bed is one of probabilistic reasoning, in which the presence or absence of benthic data provides evidence regarding the state of health of the river. The evidence is assessed or weighed in some form by the expert and integrated to reach a decision. By reformulating this decision problem as one of classification under uncertainty, the weighting and integration of this evidence can be formalised within the decision algorithm used.

Following the experiments in probability distribution adjustment it was realised that the device of evidential discount allowed finer control over the strength of evidence from a sensor or class of sensor states. For instance, evidence from sensors in the rare or absent states could be eliminated completely from evidential combination by using a discount rate of 1.0. This method also provides a means of contrasting Bayesian and Dempster-Shafer evidential combination. For singleton support functions, a discount rate of 0.0 reduces a Dempster-Shafer basic probability mass assignment to the Bayesian probability distribution. Intermediate values allowed the strength of evidence associated with particular sensor states to be varied arbitrarily, and to monitor the effect on classification performance.

The effect of varying evidential discount (i.e. strength of evidence) on classification performance was investigated using both the Bayesian and Dempster-Shafer calculus. To allow comparison with the Bayesian approach the evidence for this series of tests was represented as singleton support with varying degrees of discount ϵ depending on the sensor state. This allows the weight given to the type of evidence to be adjusted, depending on the confidence one has in the quality of the data (Dillard, 1992).

For instance, if one views evidence provided by the sensor in its abundant state as a valuable indicator of river water quality, then the discount rate for this evidence would probably be set near or equal to zero. With values of ϵ equal to zero or one, the Dempster-Shafer calculus applied to evidence represented as singleton support should reduce to the Bayesian calculus. For intermediate values of ϵ, the effect of evidential discount should lead to different results between the two calculi, since the Dempster-Shafer approach allows ϵ to represent uncommitted belief, a facility absent from the Bayesian method. These series of tests were intended to investigate the conjecture that the Dempster-Shafer method should be more robust than the Bayesian approach in dealing with the noise or imprecision inherent in the biological data, and to investigate the effect of adjusting the weight given to the sensor state evidence.

6.4.6.2 Absent Evidence

Each test consisted of a combination of discount-rates for the four sensor states. The evidence was presented to the algorithms in sensor-state order, i.e. Abundant evidence, followed by Established, Rare and Absent. In this series of experiments, Rare evidence was completely discounted, i.e. the sensor in this state was viewed as conveying no useful information regarding the quality of water at the site where it was sampled. Absent evidence was discounted at various rates between $\epsilon = 0$ and 1 to investigate the conjecture that this form of evidence, while of less significance that the three present states, was nevertheless an important factor in the performance of the automatic classification. Discount rates for evidence corresponding to the sensor states of established and abundant were kept low, i.e. this data was considered of high quality. **Table 6.9** shows data quality descriptions corresponding to the discount rates (ϵ) used.

Table 6.9 Data quality descriptions corresponding to evidential discount rates.

Data quality	Discount rate
Certain	0.0
High	0.1 - 0.2
Good	0.3 - 0.4
Fair	0.5 - 0.6
Poor	0.7 - 0.8
Uncertain/Ignore	1.0

To reduce the number of experimental variables, the area-adjusted mode of probability distribution representation was used for all the sensor evidence.

6.4.6.3 Results: Absent Evidence

Table 6.10 shows the classification rates for both calculi in which the evidence is represented as Bayesian or Singleton support. Where the evidence is Certain ($\epsilon = 0$) or Ignored ($\epsilon = 1$), singleton support reduces to Bayesian belief. Here Dempster's rule for combining evidence is equivalent to Bayes' updating rule. This is shown by the identical classification rates for the first two tests. Subsequent tests adopt a nominal discount rate for Abundant and Established evidence, to maintain non-Bayesian belief functions for the classifiers using the Dempster-Shafer algorithm.

Table 6.10 Variation of classification rates for Bayesian and Dempster-Shafer algorithms with data quality of Absent evidence.

Test ID	T221	T222	T223	T224	T225	T226	T227	T228	T229	T230
Decision Algorithm	Bayes	D-S	Bayes	D-S	Bayes	D-S	Bayes	D-S	Bayes	D-S
Classification Rate(%)	77.36	77.36	73.58	71.70	69.81	58.49	67.92	62.76	58.49	60.38
Abundant	Certain		High		High		High		High	
Established	Certain		High		High		High		High	
Rare	Ignore		Ignore		Ignore		Ignore		Ignore	
Absent	Certain		Good		Fair		Poor		Ignore	

Note: D-S = Dempster-Shafer

The decline in classification rates as Absent evidence is downgraded in quality seems to support the conjecture that this data may make an important contribution to the performance of the classification. The reasons for this are discussed below.

6.4.6.4 Discussion

The question arises however why the benthic taxa that are present in the sample do not in themselves yield good classification results over the entire data-set. **Table 6.11** shows the classification results using the Dempster-Shafer method for Abundant and Established evidence only (reference T230 in **Table 6.10**).

Table 6.11 Dempster-Shafer classification of Yorkshire data. Present data only.

		Expert Decision				
		B1a	B1b	B2	B3	B4
Classifier Output	B1a	4	2	0	0	0
	B1b	2	4	1	0	0
	B2	0	2	9	7	0
	B3	0	0	0	13	6
	B4	0	0	0	1	2

Classification rate 60.38%.

The classifier upgrades by one class from the expert's classification of B3 and B4 sites in particular. However some of these misclassifications are a result of the need to choose the maximum probability value as indicative of the decided class when constructing the confusion matrix. Examination of the performance indicators *SCI*, *I*-value and nearest system class reveal that in fact the agreement between the expert and the classifier is closer than that suggested by the base class, as used in the confusion matrix. **Table 6.12** shows the output and calculated parameters for the eight B3 sites misclassified. For site 438B, the system's intermediate class of B2-- can be interpreted as approximating the expert's opinion of B3+. For site 601B, the classifier's base class output of B4, obtained from considering the maximum value of the distribution, appears to be a clear misclassification by one class. Examination of the *SCI* and the low *I*-value reveals that this is not a clear-cut decision, however. The classifier's intermediate class of B3-- is close to the expert's B3-.

Table 6.12 Comparison of classifier output and expert decisions for eight misclassified B3 sites of test ID T230.

Site	SCI	I-value	Expert	Base Class Expert	Base Class System	System Interme-diate Class	Classifier Output				
							B1a	B1b	B2	B3	B4
ESC	2.28	13	B3+	B3	B2	B2-	0.00	0.02	0.69	0.29	0.00
435F	2.17	16	B3++	B3	B2	B2	0.00	0.00	0.82	0.17	0.00
438B	2.37	12	B3+	B3	B2	B2--	0.00	0.01	0.62	0.37	0.00
440A	2.14	16	B3+	B3	B2	B2	0.00	0.01	0.84	0.15	0.00
443A	2.17	16	B3++	B3	B2	B2	0.00	0.01	0.81	0.18	0.00
545F	2.28	11	B3+	B3	B2	B2-	0.00	0.07	0.59	0.33	0.01
601B	3.35	9	B3-	B3	B4	B3--	0.01	0.02	0.07	0.43	0.49
620A	2.24	14	B3++	B3	B2	B2-	0.00	0.01	0.73	0.26	0.00

See text for explanation.

Yorkshire Water site 440A is characterised by a benthic sample in which the presence of several taxa suggest support for higher-quality water. The presence of *Bithynia tentaculata*, *Lymnaea peregra*, *Planorbis* spp. for instance, give dominant support for class B2, with the distribution for *Sphaerium* spp. having its maximum value for B1b. It is possible that the expert gave extra weight to the presence of Tubificidae, *Erpobdella octoculata* and *Asellus aquaticus* beyond that suggested by the individual distributions themselves. The combination of these taxa together may outweigh, in the expert's mind, the contribution of the higher-quality taxa in a way that is not captured by the weighting applied for the decision algorithms.

The performance of the two calculi is more closely matched than that suggested by the classification rates over the data-set, in which classifier output is aligned with the nearest whole class. However, misclassifications also occurred for the very good quality sites (B1a). **Table 6.13** shows comparative performance of the classifiers for a particular Yorkshire site, 553B, classified by the expert as B1a, i.e. of the highest quality. The diversity of the sample was doubtless an important factor in this decision. Yorkshire Water Authority's own DoE classification of the site is class A. From **Table 6.13** it is seen that combination of evidence with increasing discount rates reduces the indicator (*I*-value) of the resulting distributions. This results in a shift from support for B1a towards B1b.

Table 6.13 Comparison of Bayesian and Dempster-Shafer classifier performance under varying data quality.

	Di	O	Di	O	Di	O	Di	O	Di	O	Di	O
B1a	0.61	1	0.57	1	0.50	1	0.52	1	0.40	2	0.48	2
B1b	0.39	2	0.41	2	0.50	2	0.46	2	0.59	1	0.48	1
B2	0.00	3	0.02	3	0.01	3	0.03	3	0.01	3	0.03	3
B3	0.00	4	0.00	4	0.00	4	0.00	4	0.00	4	0.00	4
B4	0.00	5	0.00	5	0.00	5	0.00	5	0.00	5	0.00	5
Base Class	B1a		B1a		B1b		B1a		B1b		B1b	
SCI	0.40		0.44		0.51		0.51		0.61		0.55	
I-value	12		11		9		10		11		9	
Int. Class	B1a--		B1b++		B1b++		B1b++		B1b+		B1b+	
Algorithm	Bayes		D-S		Bayes		D-S		Bayes		D-S	
Test ID	T223		T224		T225		T226		T227		T228	
Abundant	High				High				High			
Established	High				High				High			
Rare	Ignore				Ignore				Ignore			
Absent	Good				Fair				Fair			

Site: Reference 553B - River Ribble (Calder Tributary), off Washpitt Mill. Classified by the expert as B1a.
Key: D_i = support for each class; O = decision order.

Differences between the two calculi are apparent under evidential discount. With Bayesian belief, unit probability mass is redistributed across the hypotheses resulting in a smoothing of the distribution, whereas using the Dempster-Shafer mass allocation a degree of uncommitted belief equal to the discount rate ϵ is assigned to the environment Θ. In the latter case, the strength of evidence for each singleton is uniformly reduced by a factor $(1-\epsilon)$.

Table 6.14 illustrates the combination of sensor evidence at a Yorkshire site (611: Clayton Beck, an Aire tributary) for the two calculi with the same data weighting. In (a) the environment Θ has no role, since no such provision is made in Bayesian methods. In (b), uncommitted belief resides with Θ. This probability mass is available for combination with new sensor evidence. The difference in the combined support resulting from two sets of evidence is small. However, the cumulative effect from combining all the available evidence results in an appreciable difference in the final class support between the two calculi for the same data quality.

Table 6.14 Comparison of Bayesian and Dempster-Shafer evidence combination under evidential discount

Taxon	Sensor-state	B1a	B1b	B2	B3	B4	Θ
Potamopyrgus jenkinsi	ESTABLISHED	0.32	0.15	0.28	0.22	0.03	0.00
Effective Support		0.31	0.15	0.27	0.22	0.05	0.00
LUMBRICULIDAE	ESTABLISHED	0.05	0.32	0.29	0.30	0.04	0.00
Effective Support		0.06	0.31	0.28	0.29	0.06	0.00
Combined Support		0.09	0.22	0.37	0.31	0.01	0.00

(a) Bayesian updating. Test ID 225

Taxon	Sensor-state	B1a	B1b	B2	B3	B4	Θ
Potamopyrgus jenkinsi	ESTABLISHED	0.32	0.15	0.28	0.22	0.03	0.00
Effective Support		0.29	0.13	0.25	0.20	0.03	0.10
LUMBRICULIDAE	ESTABLISHED	0.05	0.32	0.29	0.30	0.04	0.00
Effective Support		0.04	0.29	0.26	0.27	0.04	0.10
Combined Support		0.12	0.22	0.33	0.28	0.02	0.03

(b) Dempster-Shafer updating. Non-zero belief remains with Θ after combination. Test ID 226

Examination of the intermediate calculations shows that the cumulative effect of absent evidence is generally to reinforce the support provided by the established and abundant sensor evidence, although it has a lower information content. Since the distributions for absent evidence span the propositions, its effect is to maintain non-zero support for propositions throughout the evidential combination procedure.

It appears that the representation of belief in both cases as singleton support is the dominant factor, rather than the calculus used, coupled with the "damping" effect of absent evidence. The cumulative effect of combining the absent evidence is to reduce the width of the evidential interval in the Dempster-Shafer case, since the size of the uncommitted belief assigned to Θ reduces each time sensor evidence is combined. Thus singleton support functions become more Bayesian as sensor evidence is successively combined using Dempster's rule.

6.4.6.5 Established evidence

Evidential strength was also varied for Established evidence. **Table 6.15** summarises the results of Dempster-Shafer classification when varying the strength of Established evidence.

Table 6.15 Classification rates for varying weighting of Established evidence.

	T230	T234	T236	T2361
Abundant	High	High	High	High
Established	High	Good	Fair	Poor
Classification rate (%)	60.38	62.26	58.49	62.26

Note: Dempster-Shafer classifiers. Absent and Rare evidence neutral.

These differences, which are not significant, suggest that the 'interference' effects obtained by combining discounted evidence mitigate any expected improvement in classifier performance due to reduced conflict.

6.4.6.6 Rare Evidence

In the series of tests described previously the role of sensor evidence in the Rare state has been assumed to be neutral, i.e. the effect of this evidence has so far been ignored. This premise is open to question, since the diversity of sample sites clearly has an influence on the expert's classification: usually, but not always, strongly suggesting that the river water quality is high. The Yorkshire data-set consists of at least fifteen sites that have six or more taxa recorded as Rare, and in only three sites are taxa there either Established or Abundant only.

During the knowledge acquisition sessions the expert was generally of the opinion that evidence from taxa occurring in very small numbers (generally ≤ 3) had to be treated with caution, since their inclusion in a sample count may be due to inconsistencies in sampling, errors in collating data, or even due to their drifting from sites upstream. Consequently, probability distributions were not elicited for taxa in the Rare state.

The effect of Rare taxa could be considered in one of two ways: the added diversity of the sample could be a weighting factor, or some representation of the evidence in this state could be adopted. In this experiment, the latter option was chosen in which Rare evidence was seen as a weaker form of Established evidence. The Dempster-Shafer algorithm is an intuitively attractive mechanism for this representation, since our uncertainty in how much weight to accord this evidence can be represented as uncommitted belief. Two evidential discount regimes were investigated: one in which Rare evidence was counted as 'fair-quality' Established evidence, the other as 'poor-quality'. Higher data quality rating was considered unjustifiable in the light of how Rare evidence was initially considered.

6.4.6.7 Results: Rare Evidence

Inclusion of Rare evidence makes an insignificant difference to the classification rates across the data-set, or to individual sample classifications, even for samples in which there were several taxa in the Rare state. **Table 6.16** shows the classification rates for three regimes of data quality. In each case Abundant and Established evidence was considered of high quality, while Absent evidence was ignored. The confusion matrices for the first and third case are identical.

Table 6.16 Classification rates for varying weighting of Rare evidence

	T230	T2381	T238
Rare	Ignore	Fair	Poor
Classification rate (%)	60.37	62.24	60.37

Note: Dempster-Shafer classifiers. Absent neutral. Established and Abundant evidence included.

6.4.6.8 Discussion: Rare Evidence

These results support the assumption that Rare evidence can effectively be ignored. From a computational viewpoint, the combination of the extra evidence makes little difference to a classifier's decision. Support for each hypothesis increases asymptotically after several combinations of evidence. Extra support for a particular class from considering Rare evidence is offset by contradictory support for competing hypotheses. These results follow from the probabilistic representation of Rare evidence as a poorer form of Established evidence. From another viewpoint, taxa that occur in low numbers could be seen as on the periphery of their ideal habitat, exhibiting a bimodal distribution of support across the quality classes, with the modes occurring either side of the preferred class. The investigation of a more accurate probabilistic representation of Rare evidence, involving the elicitation of distributions specifically for this state, could be an area for future work.

6.4.6.9 Summary: Variation in Evidential Strength

The performance of the automatic classifier under evidential discount is naturally dependent on the composition of the sensor evidence for the site. In this respect predicting the performance of an uncertain reasoning classifier over a data-set for a particular regime of evidential discount is difficult, but the effects can be understood with reference to individual sets of evidence. With the Dempster-Shafer calculus the effect of discounting evidence is complicated by the compensating role of the environment Θ, and the decrease in evidential conflict. As discount is increased for sensor evidence, the strength of that evidence decreases. For singleton support, this decrease is in direct proportion to the discount rate, while Θ receives a correspondingly increased support.

Consider the situation for incoming evidence from a sensor in the Established state, with an increased discount rate of 0.2 over the nominal 0.1. The combined evidence of previous sensors will be represented as singleton support, with non-zero probability mass assigned to Θ. The orthogonal sum of these two support functions may result in certain singletons receiving more support than that received for lower discount rates. This is because of support arising from the intersections, in Dempster's rule, between Θ and a singleton, which results in support for that singleton represented by the mass product. The reduced evidential strength for the singletons results in a lower conflict mass accumulating in the null set, and therefore a higher normalisation factor. As normalisation redistributes probability mass back to the singletons and to Θ, the lower conflict ensures that Θ continues to play a role in subsequent sensor readings of supporting all the singletons, if the evidence is represented as singleton support.

Because of these effects, there is not a straightforward relationship between varying discount rates and classification performance. Increasing discounts beyond 20% may however lead to undesirable results. The indicator values of the sensors become masked by the tendency to "even out" the support functions as more evidence is combined. If the un-discounted evidence nominally supports higher-quality classes, discounting shifts support towards the lower-quality propositions. Since larger support remains with the environment, the uncertainty in the decision, i.e. the evidential width, is larger.

For nominal discount rates for present data, the results suggest little difference between the Bayesian and Dempster-Shafer approach. However the divergence between the Bayesian and Dempster-Shafer approach increases with increasing evidential discount, since only the latter calculus has a mechanism for dealing with the uncommitted belief it represents.

6.4.7 Order of Combination and Decision thresholds

6.4.7.1 Motivation

An important characteristic of diagnostic reasoning is that decisions should be independent of the order in which evidence is presented. Therefore the automatic classifiers should reach the same conclusions for a benthic sample if evidence is ordered into sensor states (e.g. Abundant followed by Established) or if another ordering is used. However, it is possible that the expert considers benthic data in an "intuitive" order, i.e. those taxa present first, before appealing to the significance of important taxa absent from the sample. The expert may not consider every piece of evidence, but instead may decide after observing the presence or absence of the most important indicators, dismissing conflicting or weak evidence as extraneous. If this or another order is imposed, there may be no need to consider the entire set of evidence from the indicator taxa - the expert may be satisfied is the support for a particular class is say, 85% or 90%.

The decisions reached by the automatic classifiers should be identical for the same type of belief representation and evidential weight regardless of the order in which the evidence is presented, if the evidence set is the same and that the decision threshold is at or very near unity. If however an ordering is imposed, a decision threshold less than unity should produce a decision when the support for a proposition attains the threshold value, eliminating any further evidence from consideration. Such a strategy may be adopted by a human expert or by a reasoning system for a large set of evidence, in that a reduced set of significant evidence is considered. Rather than consider the entire set, the expert (or reasoning system) may reach a decision when satisfied that the important evidence has been considered. This important evidence would naturally be scrutinised first.

6.4.7.2 Procedure

To investigate this conjecture the decision threshold was varied from its default value of unity for two different orderings. The default or 'taxonomic' ordering corresponds to the order in which the indicator taxa are listed within the sample data, that is, in ascending order of Maitland codes (Maitland, 1977). The second ordering or 'sensor-state' corresponds to the "intuitive" order of considering abundant taxa first, followed by established, rare and absent taxa. If as conjectured this corresponds to the ordering imposed by the domain expert, one would expect the use of decision thresholds to be more appropriate here.

Three levels of decision thresholds were imposed: definite, high and low, corresponding to values of 1.0, 0.8 and 0.6, for the two orderings of evidential combination. The classifier reported its decision when the support value for any hypothesis reached the decision threshold.

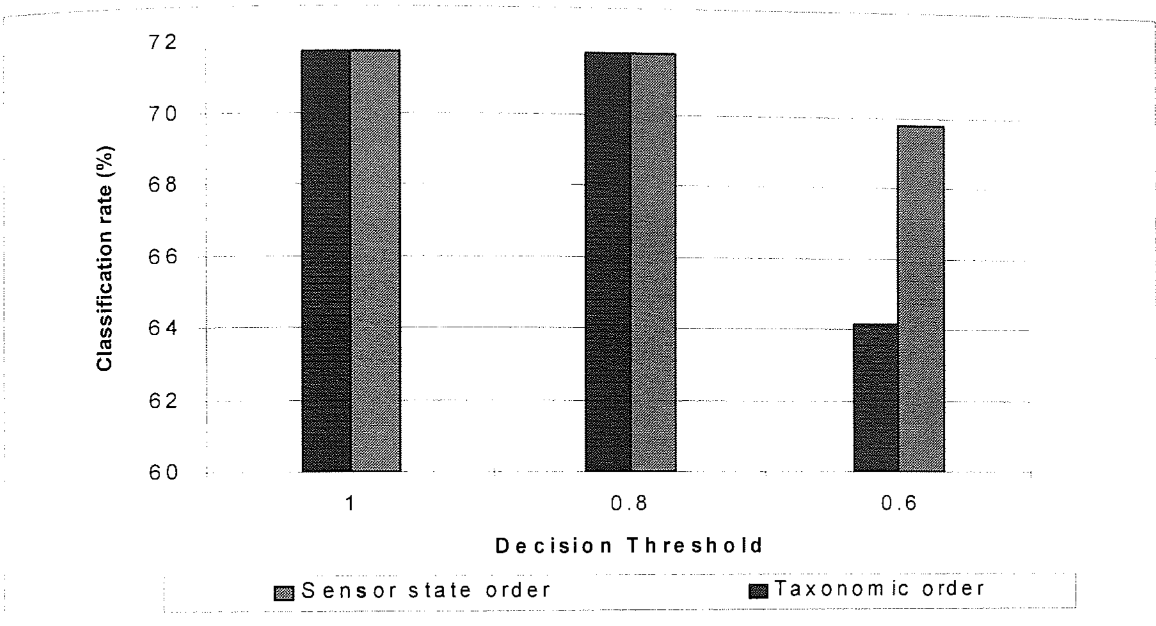

Figure 6.1 Effect of varying decision threshold on classification rate for two orderings of evidence combination. 'Taxonomic ordering' refers to sensor evidence combined in order *Polycelis nigra* to *Simulium ornatum*, regardless of sensor-state. 'Sensor-state' order means that Abundant evidence was combined first, followed by Established, etc.

6.4.7.3 Results

Figure 6.1 shows the overall classification performance across the Yorkshire data-set. As expected there is no difference between the classification rates for the taxonomic and sensor-state ordering for a decision threshold of unity. Reducing the threshold to 0.8 yields the same classification rate, although the number of indicator taxa that participate in the classification decision is reduced. The classification of a sample can therefore be obtained with a reduced evidence set by a small reduction of the decision threshold. The identical results for taxonomic and sensor-state orderings for a threshold of 0.8 are surprising in that different numbers of taxa participate in the classification decision for each sample. The differences are not significant enough to alter the 'base class' decision due to the participation of a sufficient number of high-indicator taxa for each ordering.

6.4.8 Use of Evidential Indicator Value

6.4.8.1 Motivation

Previous tests have involved weighting evidence according to a class of sensor evidence, i.e. according to sensor state (Abundant, Established, Rare or Absent). In this series of tests, the intrinsic value of each sensor as an indicator of water quality was used to determine the effect on classifier performance. A suitable value is the ratio of maximum to non-zero probability value within the distribution. For a probability distribution with zero information content (e.g. $H_i = 0.2 \; \forall \; i$) the indicator value is unity, while for strong indicators (such as abundant sensors) the indicator value is usually large. If below a threshold value, the evidence is not used in the decision process, since it is deemed to be of little value in that decision. This is equivalent to using evidential discount, but on a sensor-by-sensor basis rather than sensor state. If the sensor's evidence is below the threshold value, it is discounted entirely. The indicator value was calculated from the undiscounted probability distribution. Those distributions whose indicator value was greater or equal to a specified threshold were discounted at a nominal 0.1 to maintain non-Bayesian belief functions for the Dempster-Shafer classifier.

6.4.8.2 Results

The results confirm those found using Absent data, namely that the inclusion of low-information sensor evidence improves classifier performance over the data-set. Most Absent sensor evidence is eliminated when the indicator value is greater or equal to 10, along with some Established evidence. **Table 6.17** shows the classification performance using those sensors with indicator-values greater or equal to 30. The misclassifications of the B3 sites are due to the occurrence in those samples of taxa associated with B2 classes. For instance the Yorkshire site 435F classified by the expert as B3 has *Lymnaea peregra*, *Ancylus fluviatilis* and *Sphaerium* spp. in the Abundant state, giving strong support for class B2, the conclusion arrived at by the classifier.

Table 6.17 Effect of counting only sensor evidence greater or equal to given 'Indicator value'

	T224	T286_I05	T286_I10	T286_I20	T286_I30
Indicator Value	1	5	10	20	30
Classification Rate %	71.70	67.92	60.38	58.49	45.28

See text for explanation.

In one case, a B3 site (494) was seriously mis-classified as B1a. The fauna at this site is sparse, consisting only of the indicator taxa *Ancylus fluviatilis*, Tubificidae and *Erpobdella octoculata*, all in the Established state. A high indicator value threshold of 30 eliminated Tubificidae and *Erpobdella octoculata* from consideration resulting in a marginal decision from *Ancylus fluviatilis* for class B1a. The performance over the entire data-set is shown in **Table 6.18**.

Table 6.18 Classification of Yorkshire data using indicator value of 30

		Expert Decision				
		B1a	B1b	B2	B3	B4
Classifier Output	B1a	3	2	1	1	0
	B1b	3	6	3	2	0
	B2	0	0	6	9	0
	B3	0	0	0	4	3
	B4	0	0	0	5	5

Classification rate 45.28%.

The inclusion of strong indicators only results in premature decisions arrived at by a subset of the available evidence. As with evidence from absent taxa, including evidence with lower-information content improves classification performance for the multihypothesis belief representation of singleton support.

6.4.9 Conflict resolution for singleton support functions

6.4.9.1 Background

This section deals with experiments that attempted to monitor and resolve conflict that occurs during the accumulation of evidence for and against the propositions corresponding to the biological quality of river water. It begins with a reexamination of conflict within the Dempster-Shafer calculus and its monitoring within various computational schemes. The implication of the choice of a scheme for the order in which evidence is presented is then discussed, before considering the results of the experiments themselves.

Belief represented as singleton support can be combined using one of two schemes.

For pair-wise applications of Dempster's rule between current evidence and incoming sensor data, any conflict arising from this operation is visible as mass in the null-set, which is reset to zero by normalising the probability masses assigned to the focal subsets of Θ *before* the next sensor's data is combined. In using Barnett's scheme, normalisation takes place after the total evidence for and against the singleton propositions is accumulated.

One clear advantage in using Barnett's scheme is the ability to monitor the total weight of conflict in combining the sensor data continuously. This total weight of conflict is lost in using pair-wise applications of Dempster's rule, in which the process of normalisation redistributes the conflicting mass back to the focal elements between each combination. The conflict in the pair-wise case is that between the current combined evidence and the incoming sensor data. If in combining the new evidence the mass of the null-set is large, a reasoning system can reject it and maintain the support for the propositions at the state before the incoming data. This rejection of the sensor data can however be problematical. For very high conflict (i.e. $m(\oslash) \cong 1$), the rejection of the conflicting data is obligatory, since the orthogonal sum then will not be defined. For values lower than unity the decision to accept or reject the new sensor data can become arbitrary unless one has further knowledge on the sensor evidence.

One viewpoint may be to consider the evidence as erroneous, like a malfunctioning electronic sensor, or anomalous, such as the unexpected appearance of taxa normally associated with poor quality waters in good quality classes. Another viewpoint is to consider such anomalies as inevitable in freshwater communities subject to stochastic events. If this viewpoint is adopted, the appearance of certain 'conflicting' taxa should perhaps be accommodated by the reasoning scheme rather than rejected automatically.[3]

Since the Dempster-Shafer classifiers for singleton support functions were implemented using the pair-wise application of Dempster's rule, only the evidential conflict between existing and incoming sensor data was monitored. The actual degree of conflict depends on the current belief with the focal propositions and the incoming data, whose occurrence is dictated both by the ordering scheme and the degree of evidential discount. For $\epsilon = 1$, the combination will not occur, otherwise the basic probability assignment corresponding to the sensor evidence will be combined.

[3] A procedure for dealing with evidential conflict in biological classification for the Bayesian classifier has been described in Walley *et al.* (1992a), in which the sample data is checked for conformity. Potentially contentious evidence whose inclusion would unduly influence the classification is identified and if necessary rejected.

At a particular moment in time during the combination procedure the conflict between the current evidence and the incoming sensor data will depend on the order in which that data is presented. Taxonomic order of sensor input may exaggerate the conflict if, for instance, evidence associated with sensors in the absent state is accumulated first, followed by established or abundant evidence. It is quite possible that incoming sensor evidence corresponding to present taxa may be highly conflicting with current evidence accumulated by observing absent taxa; in which case there is a danger of rejecting incoming sensor data that would otherwise be accepted. Therefore ordering the sensor evidence seems intuitively more satisfactory so that present data are considered before absent taxa. Any conflict arising between present taxa then has more significance, and is less likely to result in poor decisions to reject sensor evidence.

6.4.9.2 An example of conflict monitoring and normalisation

The following example will illustrate the mechanism of conflict resolution using Dempster's rule, and show how empirical threshold values can be used to manage conflict. It begins by examining evidential combination in which the conflict threshold is set to unity, i.e. all evidence except that which is completely contradictory is combined with existing support for the propositions. The example shows the intermediate calculations for the classification of site 610A. From the list of indicator taxa, the domain expert assessed this site to be of class B1a, i.e. very good quality.

Evidence is presented for combination in sensor state order, so that abundant evidence is considered before established evidence. In this example rare and absent evidence has neutral effect. For each set of evidence, the corresponding discrete probability distribution for the sensor state is read from a benthic database and presented for combination. **Table 6.19** gives the sensor state and the corresponding discrete probability distribution across the five quality classes, from which is derived the singleton support function.

The first row in the table shows the combined support of previous evidence (not shown) from *Potamopyrgus jenkinsi* and *Pisidium* spp. The freshwater shrimp *Gammarus pulex* in the Established state is considered by the domain expert to be a strong indicator of B1a quality: this is reflected in the values of $P(H_i|e)$ for this sensor state. These values are then used to make the basic probability assignment across the singletons. Here the effective support for each singleton is reduced by a factor $(1 - \epsilon)$, where $\epsilon = 0.1$, assigned as a basic probability number to Θ. This "effective support" is shown in the row following the probability assignment for *Gammarus pulex*.

Table 6.19 Evidential combination and conflict resolution in Dempster-Shafer calculus

State of evidence	B1a	B1b	B2	B3	B4	Θ	∅
Combined support	0.34	0.15	0.28	0.18	0.01	0.025	0.00
Gammarus pulex	0.44	0.22	0.29	0.04	0.00	0.00	0.00
Effective support	0.39	0.20	0.26	0.04	0.00	0.10	0.00
Combined support (unnormalised)	0.18	0.04	0.10	0.02	0.001	0.002	0.63
Combined support (normalised)	0.48	0.13	0.29	0.07	0.004	0.006	0.00

This retention of uncommitted belief maintains non-zero belief in each singleton between combinations of evidence. Although *Gammarus pulex* offers zero support for class B4, the resultant probability mass after combination and normalisation is non-zero, although very small. This is important in that subsequent sensor evidence may potentially offer strong support for B4. If the probability mass for this proposition is non-zero, this subsequent sensor evidence will have an effect. If the support for the proposition had gone to zero, the subsequent sensor data would have no effect, precluding their influence on the classification decision. The sensor data is effectively eliminated, a situation that may be undesirable.

The previous evidence of *Potamopyrgus jenkinsi* and *Pisidium* spp. in the Established state are supportive of class B1a. The combined effect is to reinforce support for B1b. However because *Leuctra* spp. is strongly supportive of B1a, the combination leads to conflict, resulting in mass accumulating in the null set. Normalisation redistributes this mass back to the focal elements, including the environment Θ. The process of normalisation also prevents the probability numbers from becoming too small, which condition can cause computational problems. Note that conflict arises even with supportive evidence, to the degree of 0.63. Conflicting mass is removed from the null set by normalisation.

6.4.9.3 Rejection of evidence due to excessive conflict

Table 6.20 shows part of a calculation involving conflicting evidence. Combination of evidence provided by Ceratopogonidae, established at this site, results in an unchanged degree of support for the classes, with dominant support for B1a (0.68). The evidence provided by *Simulium ornatum* however results in a large degree of conflict (0.88), here exceeding the conflict threshold set for this classifier. The evidence is therefore rejected as inconsistent with the combined support of the previous evidence. For this sample, the overall decision reached by the classifier when considering only those taxa present was as that shown as 'Remaining support' in the table. Had the conflicting evidence been included, the decision would have been (0.58,0.42,0.00,0.00,0.00). Although the rejection of the conflicting evidence has led to a 'better' overall decision, the difference is small, and in fact in both cases the classifier outputs align on the same base class of B1a. Nevertheless, the rejection of the indicator associated with poorer-quality waters seems in keeping with the expert's opinion of a very good quality sample.

Table 6.20 Evidential combination for site 636A by Dempster-Shafer classifier

State of evidence	B1a	B1b	B2	B3	B4	Θ	⌀
Combined support	0.68	0.31	0.001	0.00	0.00	0.00	0.00
Ceratopogonidae	0.29	0.29	0.29	0.10	0.01	0.00	0.00
Effective support	0.26	0.26	0.26	0.09	0.009	0.10	0.00
Combined support (unnormalised)	0.24	0.11	0.00	0.00	0.00	0.00	0.636
Combined support (normalised)	0.68	0.31	0.00	0.00	0.00	0.00	0.00
Simulium ornatum	0.00	0.065	0.45	0.425	0.06	0.00	0.00
Effective support	0.00	0.058	0.405	0.382	0.054	0.10	0.00
Combined support (unnormalised)	-	-	-	-	-	-	0.88
Remaining support	0.68	0.31	0.00	0.00	0.00	0.00	0.00

Note: Site classified by the domain expert as B1a. The classifier used a conflict threshold of 0.8, leading to the rejection of evidence from *Simulium ornatum*.

The classification performance over the data-set shows that marginal reductions in conflict thresholds have, like varying decision thresholds, little effect, suggesting that high levels of conflict are the norm rather than the exception. This follows from the representation of benthic data as singleton support or Bayesian belief. Reducing the conflict threshold to 0.6 results in a marked increase in the number of taxa rejected, leading to poor classifier decisions (**Table 6.21**).

Table 6.21 Effect of enforcing conflict threshold on classification performance

	T241	T257H	T257M	T258H	T258M
Conflict Threshold	Inf.	High	Medium	High	Medium
Combination Order	T	T	T	S	S
Classification Rate %	71.70	67.92	47.17	71.70	54.71

Notes: Conflict thresholds vary from Infinite (1.0) to High (0.8) and Medium (0.6). The evidence was presented to the Dempster-Shafer classifier in both Taxonomic (T) and Sensor-state (S) order.

6.4.9.4 Discussion: Conflict thresholds

There is suggestive evidence that sensor-state combination yields better classification rates than taxonomic-order combination when a moderate level of conflict threshold is employed using the Dempster-Shafer algorithm, lending support to the conjecture that this models the combination order used by the domain expert. Dempster's rule deals with highly conflicting evidence by promoting the overall consensus. However, the order of combination is important. Since the degree of evidential conflict, as measured by the mass of the null-set, is a dynamic quantity that is dependant on the previously combined evidence and the incoming sensor evidence, the conflict level is highly dependent on the particular sample data. (This follows from the use of pair-wise combinations in Dempster's rule). Altering the order of combination will produce different dynamic conflict values for each orthogonal sum, and therefore the imposition of the 'natural' (sensor-state) order of combination should theoretically lead to better rejection decisions than using the taxonomic order.

For a singleton (or Bayesian) belief representation, the degree of evidential conflict will inevitably be significant, since simultaneous support for mutually exclusive propositions is conflicting *de facto*. The Dempster-Shafer calculus, unlike the Bayesian approach, allows true evidential discount by allowing for uncommitted belief. This is a mechanism for sustaining support for the propositions, controlling the strength of the evidence and consequently reducing conflict. However the cumulative effect of combining a large evidence set is to reduce the role of Θ and render the belief functions more Bayesian. Thus, in evidential reasoning the total degree of conflict increases with the size of the evidence set. (This will be apparent in the section on simple support functions in which the total weight of conflict can be monitored).

On introducing absent evidence, the degree of evidential conflict can become quite high, in spite of the lower information content. Since the probability distributions span the singletons, the mass products obtained during application of Dempster's rule will involve $n \times (n - 1) = 20$ non-intersecting sets, all equivalent to the null set. However, the inclusion of absent evidence does appear to improve classification performance. Thus the occurrence of conflict should not necessarily imply that sensor data is faulty or should be rejected. The Dempster-Shafer calculus promotes a consensus by assigning mass only to intersecting sets via the device of normalisation. So the calculus manages conflict naturally, whereas the use of empirical conflict threshold levels may distort this ability to integrate disparate and noisy sensor data.

If a conflict threshold level is used, it is suggested that it should be set very high for preventing the combination of totally conflicting evidence. The role of conflict is further examined in the experiments using simple support functions in Chapter 7.

6.5 Summary

This chapter has described part of the experimental programme conducted to investigate the biological classification of river water quality using the Bayesian and Dempster-Shafer calculi. Benthic data from rivers is represented as uncertain evidence supporting the hypotheses corresponding to these classes, and combined to decide the river quality at a riffle site. Mechanisms for arriving at these decisions, including the problem of intermediate classifications and assessing the "strength" of classifier output, have been described in detail. The role of the contingency table or confusion matrix to assess the performance of a classification regime over the Yorkshire Water data-set has been discussed.

The experiments described the representation and implementation of evidence as Bayesian and singleton-support, in which multiple singleton hypotheses are supported simultaneously by uncertain evidence corresponding to the state of indicator benthic taxa. Classification rates for both the Bayesian and Dempster-Shafer improved using the adjusted probability distributions, a device that reduces the likelihood of premature decisions from highly-focused evidence.

A series of experiments was described in which variations in data quality were applied to emulate the expert's own conjectured weighting of benthic evidence. For the multihypothesis belief representation, the inclusion of absent evidence makes a significant difference to classification rates. In spite of its lower information content, the cumulative effect of evidence induced by the absence of indicator taxa is to reinforce evidence from abundant and established taxa, while simultaneously maintaining positive support to propositions that would otherwise be vetoed during evidential combination.

The modelling of rare evidence as having neutral effect on the decision was borne out by the results. However, this may follow directly from the probabilistic representation used. It is suggested that a more intensive knowledge elicitation exercise specifically for rare taxa may lead to a more realistic representation of their effect on the direct interpretation of biological quality.

Classification performance, when measured over the data-set, was not affected by a small reduction in the decision threshold, and was identical for two different combination orderings. The use of an "evidential indicator" value provided a data-quality index intrinsic to each sensor's evidence. Results confirmed observations that the inclusion of evidence with low-information improves classification performance.

Finally, the computational behaviour of the Dempster-Shafer calculus was investigated with respect to its ability to handle evidential conflict. The multihypothesis representation of Bayesian belief and singleton support lead to a high level of conflict, which is dealt with effectively by Dempster's rule of combination. Since only totally contradictory evidence may not be combined, conflict-threshold levels may be set very high.

Chapter 7

Classification Experiments - II

7.1 Introduction

The purpose of this chapter is to discuss computational aspects and classification performance for the simple support and consonant belief functions, two classes of belief function used in the Dempster-Shafer calculus which differ considerably in form and behaviour to the Bayesian or singleton support functions. The motivation for adopting particular classes of belief function is to constrain the number of possible ways in which probability mass can be assigned within the Dempster-Shafer calculus. This is particularly true of simple support, in which mass assignments are made to singleton sets and their complements. As in Chapter 6, the calculations of classifications for individual sites are presented as a means of understanding the behaviour of the Dempster-Shafer calculus for particular belief representations.

7.2 Simple Support Experiments

7.2.1 Theoretical Background

Gordon and Shortliffe (1985) in their examination of Dempster-Shafer theory concluded the scheme was well suited to the modular nature of the rule-based MYCIN system. If the conclusion of a rule is confirmed with strength s, the effect on belief in the subsets of Θ can be represented by a *bpa* with support s focused on a singleton subset A, the remaining $1 - s$ assigned as uncommitted belief to Θ. If the rule disconfirms the singleton, the support is focused on $\neg A$, the complement of A, with the remaining assigned to Θ as before. A belief function which assigns mass to singleton subsets or their negations is known as a *simple support* function and is probably the simplest representation of evidence (see **Figure 7.1**).

If moreover the subset is a singleton or its negation, evidential combination may be carried out in linear time using a computational scheme developed by Barnett (1981) and employed by Gordon and Shortliffe (1985) for the MYCIN system. Since this scheme was used to classify the Yorkshire data-set, a brief summary is presented here. A more detailed derivation of the formulae is given in Appendix A4.

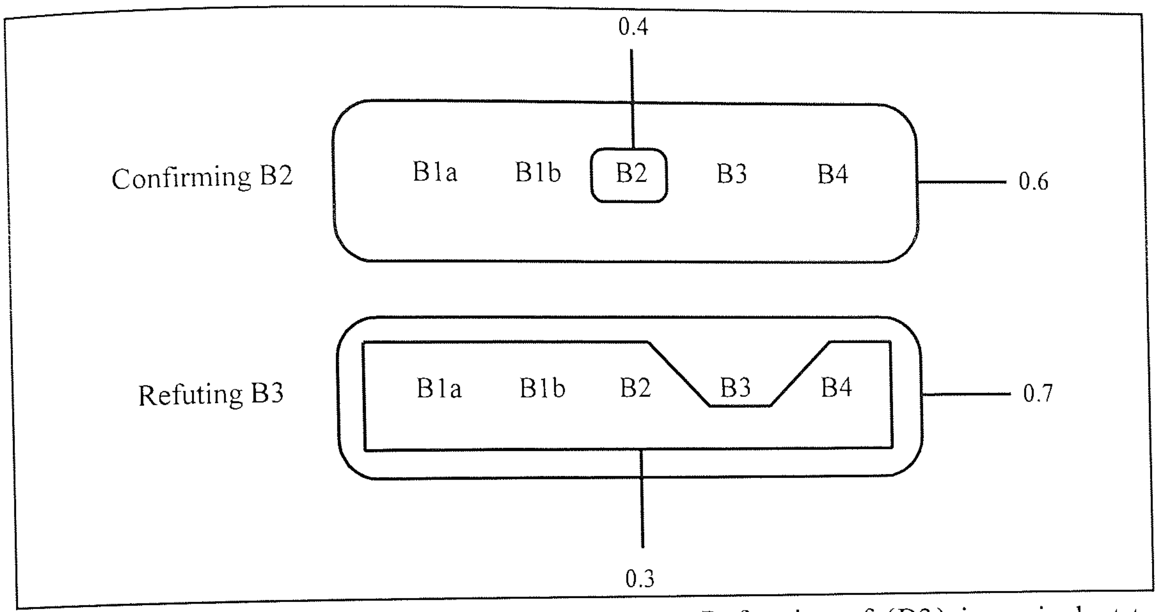

Figure 7.1 Graphical depiction of simple support. Refutation of {B3} is equivalent to support for ¬{B3}, i.e. {B1a,B1b,B2,B4}

7.2.2 Barnett's Scheme

Consider the problem of accumulating uncertain evidence for and against a particular singleton hypothesis H_i in Θ. Assume that m_1, m_2, ... m_k correspond to single support functions each supporting H_i, while m_{k+1}, m_{k+2}, ... m_{k+l} correspond to simple support functions supporting $\neg H_i$, the complement of H_i. Recall that if the degree of support is s, this is assigned to the singleton hypothesis, with the remainder 1 - s to the environment. If the evidence is against H_i, this is equivalent to support for its negation, in which case s is assigned to $\neg H_i$ and the remainder once again left as uncommitted belief with Θ. Thus there are two types of uncertain evidence in this scheme: one in favour of a particular H_i, and one against. In a practical application some of the sensors will give support to the singletons, while others will weigh against the singletons. When accumulating evidence for and against a singleton, a particular sensor cannot simultaneously support it and its complement, by the definition of the simple support function. In this discourse, 1 to k sensors are providing evidence supporting a singleton, k+1 to k+l are providing evidence against it.

Thus m_1, m_2 ... m_k represent evidence in favour of H_i so that

$$f_i = m_{for}(H_i) = m_1 \oplus m_2 \oplus \cdots \oplus m_k \tag{7.1}$$

while the total weight of evidence against H_i is given by

$$a_i = m_{against}(H_i) = m_{k+1} \oplus m_{k+2} \oplus \cdots \oplus m_{k+l} \tag{7.2}$$

The subscripts refer to each hypothesis H_i, so that there are n of these 'for' and 'against' support functions for each of the $i = 1 \ldots n$ singleton hypotheses in Θ. The combination of the for and against support functions for each of the H_i yield the simple evidence functions for the n singletons:

$$e_i = f_i \oplus a_i = m_1 \oplus m_2 \oplus \cdots \oplus m_k \oplus m_{k+1} \oplus m_{k+2} \oplus \cdots \oplus m_{k+l} \qquad \textbf{(7.3)}$$

Combination of evidence often leads to conflict. Defining a factor as $K_i = (1 - a_i f_i)^{-1}$ which accounts for this conflict we can define factors p_i as the measure of support for H_i, c_i the support against H_i, and r_i is the uncommitted belief (i.e. the support for Θ), such that $p_i = e_i(\{H_i\}) = K_i f_i(1 - a_i)$, $c_i = e_i(\{\neg H_i\}) = K_i a_i (1 - f_i)$ and $r_i = K_i (1 - a_i)(1 - f_i)$. Thus $p_i + c_i + r_i = 1$.

A simple example may illustrate these ideas. Consider a scenario in which a set of evidence provides support for the singleton A, whilst another set supports singleton B. If these are the only two singletons in the frame of discernment, support for B corresponds to evidence against A. Note that the converse is true: from B's perspective the second evidence set supports it, while the first weighs against it. Thus each sensor reading impacts on both singletons. Let us focus on the belief in A. Consider now that the evidence in favour of A is $f_A = m_{for} = 0.408$, that against A (due to the support for B) is $a_A = m_{against} = 0.22$. Clearly the two sets of evidence are conflicting, so that the mass of the null set arising their combination is

$$m_{for} \oplus m_{against} = f_A a_A = 0.408 \times 0.22 = 0.089 \qquad \textbf{(7.4)}$$

Since the mass of the null set must be zero by definition of a *bpa*, the combined evidence must be normalised by a factor $1 / (1 - af)$, i.e. the factor K defined above. Substituting in the values for f and a, we have $K = 1.098$. Therefore the measure of support for A, $p_i = K_i f_i(1 - a_i) = 0.35$. The form of this equation is seen to combine evidence for and against A while also taking into account the evidential conflict. Continuing, the measure of evidence against A, $c_i = K_i a_i (1 - f_i) = 0.14$, while the so-called residue or uncommitted belief (neither for or against A) $r_i = K_i (1 - a_i)(1 - f_i) = 0.51$. Since the combined evidence is normalised, $p_i + c_i + r_i = 0.35 + 0.15 + 0.51 = 1$. Note that this method is equivalent to using the 'intersection table' method illustrated in Chapter 3.

Generally there will be more than two singleton hypotheses ($n > 2$) so that the size of the complementary subset $\neg H_i$ is $n - 1$, where n is the size of Θ. When different H_i are supported, the effect of conflicting evidence becomes more difficult to follow, since support for the $\neg H_i$ overlap previously supported singletons. Barnett's scheme allows evidence from disparate sources to be rapidly combined, provided that this evidence is in the form of simple support. From the evidence functions overall belief functions for each hypothesis and its complement can be evaluated, thereby allowing the computation of belief intervals for each singleton.

7.2.3 Procedure

7.2.3.1 Implementation of Simple Support

In the Dempster-Shafer calculus the belief representation referred to as singleton support is essentially Bayesian belief with a degree of uncommitted belief. In contrast for simple support, belief is focused on one particular subset of Θ, either a single hypothesis or its complement. Representation of sensor evidence in this manner allows the use of Barnett's scheme for rapidly combining evidence. For its use in the biological classification scheme, there is however the problem of mapping the probability distributions which may span the range of classes to a single focus. The approach taken was, where possible, to use the maximum value within the probability distribution as the basic probability number whose focus is the associated singleton.

For the three-singleton example, consider a discrete probability distribution from sensor evidence e $\{P(A|e) = 0.2, P(B|e) = 0.6, P(C|e) = 0.2\}$. Using the distribution as a guide to making a simple support assignment, a basic probability number m = 0.6 is assigned to the proposition B, the remainder assigned to the environment Θ. For such an assignment, the main thrust of the sensor evidence is considered to lie with the proposition that has maximum support within the distribution. In this application of simple support for present taxa, the support function is said to <u>confirm</u> a particular singleton hypothesis. Note that the uncommitted belief can contribute to the plausibility of the other two singletons, so that the belief for these suggested by the distribution is not discarded. This scheme can be considered to be a "direct" use of simple support as a belief representation.

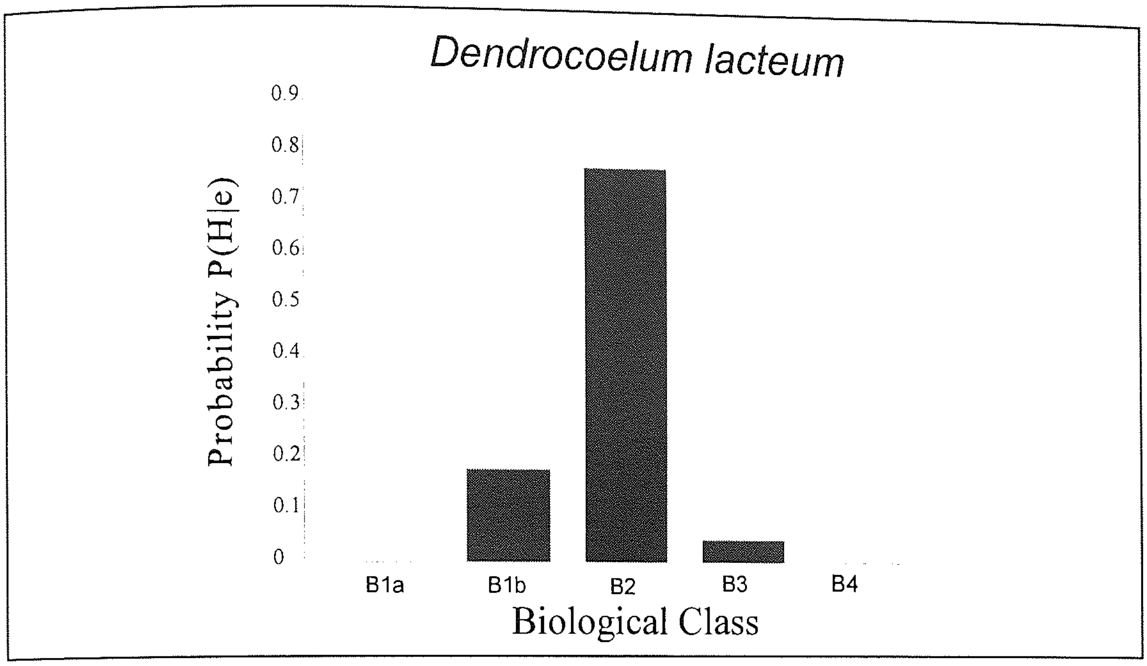

Figure 7.2 Derived probability distribution for *Dendrocoelum lacteum* in abundant state

Consider for instance the evidence provided by the observation that the sensor *Dendrocoelum lacteum* (a flatworm) is abundant at a sampled site. The derived probability distribution for the sensor in this state is shown in **Figure 7.2**, in which it is clear that this evidence provides strong support for quality class B2 ($P(H|e) = 0.78$). One obvious way to directly represent this evidence as simple support is to assign the basic probability number 0.78 to the focus {B2}, with the remaining 0.22 assigned to Θ as uncommitted belief. Thus this evidence confirms {B2} to the degree given by the maximum value in its probability distribution.

This method appears intuitively satisfactory for evidence which has "strong indicator" value, i.e. the support as suggested by the probability distribution is concentrated on a particular class. If the distribution is less sharply focused, representing the evidence in this manner is more problematic, particularly if the difference in probability mass values between the distribution maximum and its nearest competitor is small, since support for the 'competitor' class is re-assigned to Θ. For ubiquitous taxa, i.e. those which give equal support to two or more classes (for example, Lumbriculidae) the choice of the focus for simple support is somewhat arbitrary. One approach, adopted for the Dempster-Shafer simple-support classifiers, was simply to choose the median class in the distribution for ubiquitous taxa as representative of the evidence.

7.2.4 Using present data only

Evidence from indicator sensors which were present in the Yorkshire Water biological samples was directly represented as simple support, using the maximum values from the probability distributions. In general, most of the 41 sensors in the established or abundant states have sufficiently pronounced profiles to make the choice of the focus straightforward. The belief combination for the classifications were carried out using Barnett's scheme, which allows the total weight of evidential conflict to be monitored, in order of sensor-state (i.e. Abundant evidence, followed by Established).

7.2.4.1 Results

Table 7.1 shows the classification performance over the Yorkshire data-set, when the evidence is represented as simple support. The poor classification rate is due to 24 misclassifications, although all were within one-class of the expert's decision. However, nine B3 sites were upgraded by the Dempster-Shafer classifier to B2. The reasons for this are similar to those which obtained for singleton support, in that the evidence from several taxa did suggest strong support for class B2. The classification rate may be compared with an equivalent classification using evidence represented as singleton support, namely 60.38%.

Table 7.1 Classification of Yorkshire data, with evidence represented as simple support

		Expert Decision				
		B1a	B1b	B2	B3	B4
Classifier Output	B1a	4	2	0	0	0
	B1b	2	4	1	0	0
	B2	0	2	7	9	0
	B3	0	0	2	9	3
	B4	0	0	0	3	5

Notes: Only present data were considered. Classification rate 54.72%.

One problem that can arise with the Dempster-Shafer calculus is the fact that a decision may not be point-valued, i.e. the degree of support for an hypothesis is expressed as the interval [Belief, Plausibility] which in general has non-zero width. As more evidence expressed as singleton support is combined, the resulting basic probability assignment becomes increasingly Bayesian as the evidential width (the degree of belief residing with the environment) decreases. For simple support, the role of Θ is enhanced by virtue of the fact that a single focus receives a basic probability number. Because of this, the final evidential width for simple support representation is more pronounced than for singleton support, particularly with few participating taxa.

Figure 7.3 shows the variation of the degree of uncertainty (i.e the width of the evidential interval) with the number of participating taxa for this test. Using Dempster's rule, the width of the interval will decrease with each evidential combination, even with conflicting evidence, an attribute which some workers consider undesirable (Chang and Kashyap, 1990). As the number of combinations increases, the width will tend to zero, approaching the Bayesian situation of a point decision. Conversely, if only a small number of taxa have participated in the decision, the width of the interval will tend to be large.

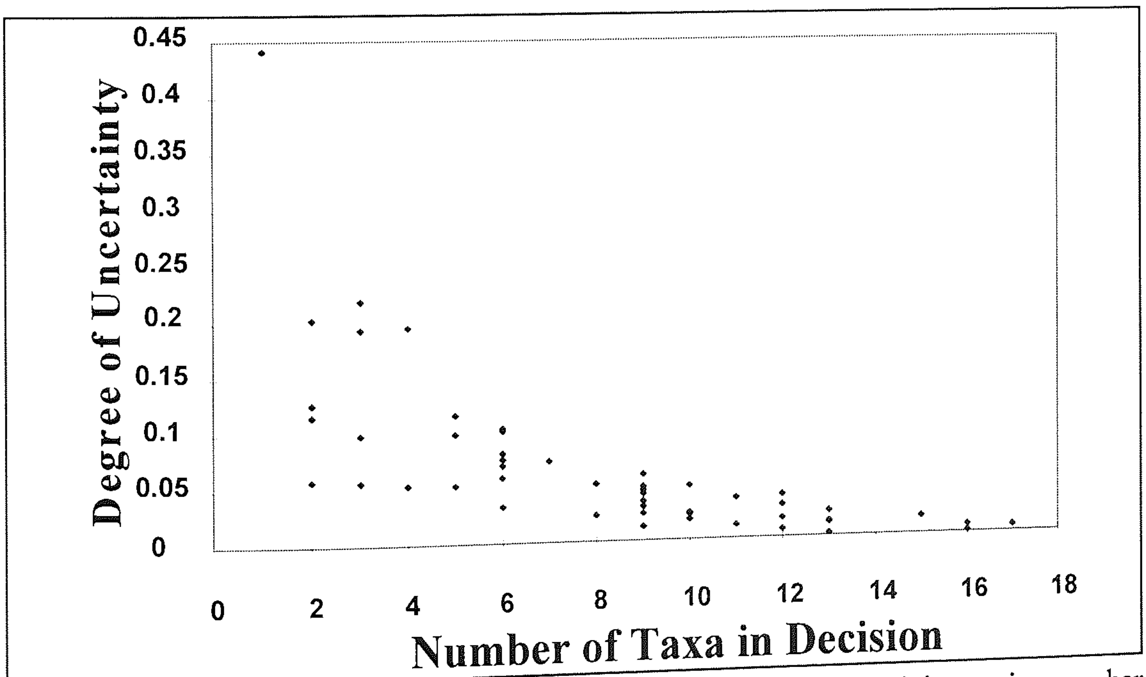

Figure 7.3 Evidential interval width (uncertainty) tends to decrease with increasing number of participating sensors

For confirming evidence (as used here), the width of the evidential interval is the same for all five classes.[1] Barnett's scheme for combining simple support calculates the support for and against the singletons. If the evidence is all confirming, any conflicting evidence for a singleton is embodied in the support for opposing hypotheses. Effectively, the direct evidence opposing a particular singleton is zero.

Table 7.2 illustrates the result of combining the evidence provided by the Yorkshire Water sample 619B. Seventeen indicator taxa are present in the sample, although five of them are in the rare state and have a neutral effect on the decision: thus there are 12 sensors participating in the decision. The evidence in favour f_i of each hypothesis H_i in calculated from equation **(7.1)**, in which each measure of support s_k (k = 1 to 12) is assigned on the basis of the probability distribution provided by the sensor state. Using the simple support assignment, the support s_k will be non-zero for hypothesis H_i if that hypothesis has the maximum probability value within the distribution. For instance the mayfly *Baetis rhodani* in the abundant state will support B1b to degree 0.525, while the other singletons receive zero support. The remaining 0.475 is assigned to the environment Θ.

[1] A proof of this is given in Appendix A4 on Barnett's scheme.

Table 7.2 Dempster-Shafer simple support classification of benthic data from Yorkshire site 619B.

EVIDENCE			
Class	For (f)	Against (a)	Conflict (K_i)
B1a	0.806794	0	1
B1b	0.813189	0	1
B2	0.908978	0	1
B3	0.873466	0	1
B4	0	0	1

(a) Evidence for, against, and degree of conflict for each hypothesis (class)

SUPPORT			
Class	For (p)	Against (c)	Residue (r)
B1a	0.806794	0	0.193206
B1b	0.813189	0	0.186811
B2	0.908978	0	0.091022
B3	0.873466	0	0.126534
B4	0	0	1

(b) Basic probability numbers arising from combining evidence in (a).

INTERNAL/OVERALL CONFLICT				
σ	K	$K.\Pi\ K_i$	$log\ K$	$log\ (K.\ \Pi\ K_i)$
26.4182	91.0584	91.0584	1.9618	1.9618

(c) Conflict values and normalisation factors

BELIEF REPRESENTATION				
Class	Belief	Plausibility	Doubt	Uncertainty
B1a	0.158067	0.195919	0.804081	0.037853
B1b	0.164773	0.202626	0.797374	0.037853
B2	0.378009	0.415862	0.584138	0.037853
B3	0.261298	0.299151	0.700849	0.037853
B4	0	0 . 0 37853	0.962147	0.037853

(d) Resulting belief function parameters after combination of all present evidence

Note: Only taxa which were present in the sample participated in the decision.

The normalisation factor K_i for each hypothesis is 1, since the evidence a_i against each hypothesis is zero. Thus the measures of support for and against each hypothesis p_i and c_i are identically equal to f_i and a_i respectively (**Table 7.2**). The residue r_i associated with the environment Θ for each hypothesis H_i is calculated from the fact that $p_i + c_i + r_i = 1$. These values therefore represent a basic probability assignment for each hypothesis, obtained by pooling the confirming (f_i) and disconfirming (a_i) evidence. These $n = 5$ basic probability assignments must now themselves be combined. The details of this combination are described in Appendix A4 which presents Barnett's scheme in detail. During the combination a factor K is calculated, which is the measure of the conflict between the singleton hypotheses.

The total weight of conflict in the combination is $\log(K. \Pi K_i)$ (Barnett, 1981), and

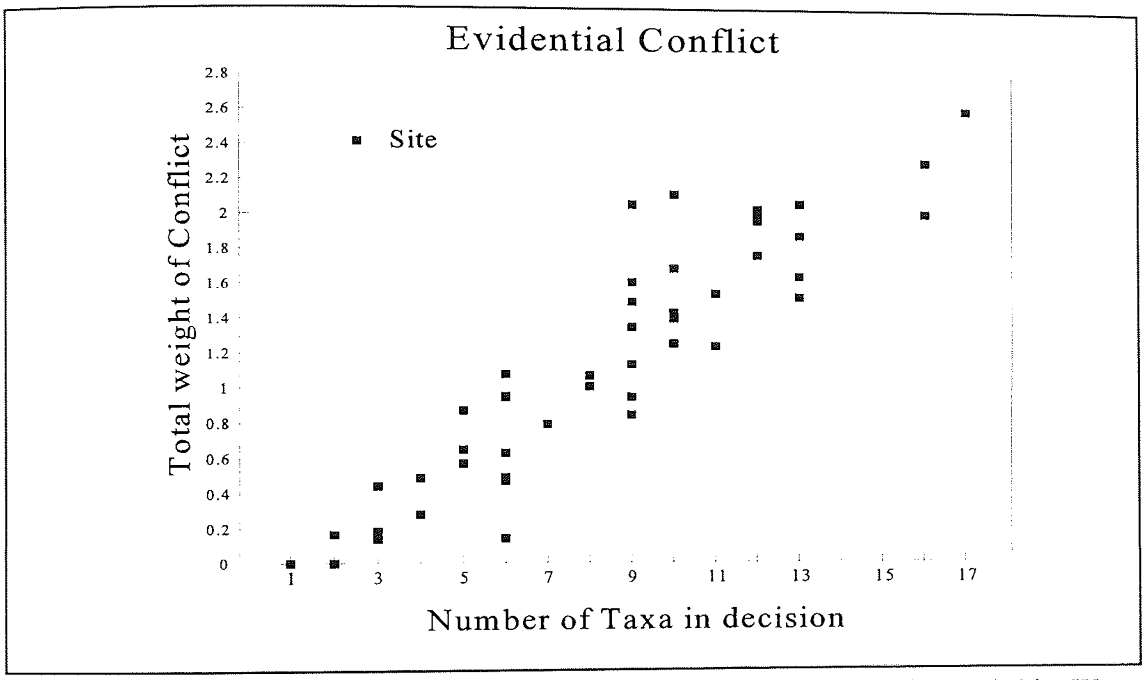

Figure 7.4 Total weight of conflict versus number of participating taxa for Yorkshire Water Data set. Simple support functions, present data only.

tends to increase with increasing numbers of participating sensors. **Figure 7.4** shows the variation of the total weight of conflict against numbers of participating taxa for the Yorkshire Water data-set. Although the uncertainty in the decision decreases with increasing sensors, the conflict increases with the number of sensors.

7.2.5 Role of Absent Evidence

7.2.5.1 Representation of evidence

The role of simple support evidence from absent indicator taxa was investigated in which the focus is a singleton hypothesis. For evidence provided by the *absence* of the sensor taxa, two approaches suggest themselves for mass allocation under simple support. The first is as before, with the maximum value in the distribution being assigned to the corresponding focus. This however is also problematic since absent evidence has much less *indicator value* than that provided by the established or abundant sensor states. **Figure 7.5** shows the probability distribution for *Dendrocoelum lacteum* in the absent state. The distribution clearly has a low information content. This follows from the fact that the probability of this taxon being absent in any of the classes is high, and therefore its absence signifies little about the quality of the river water.

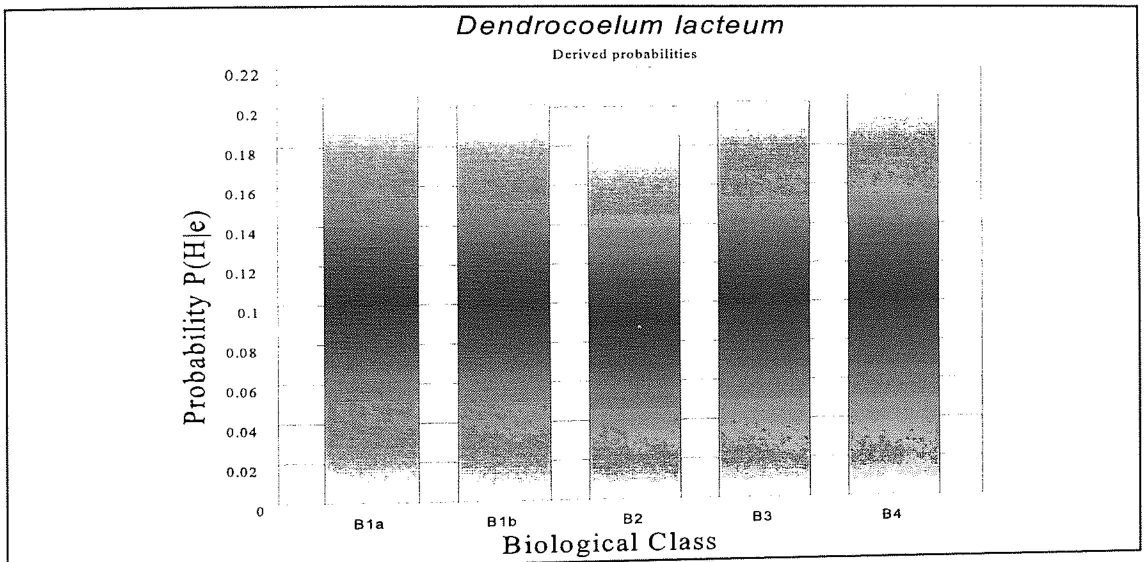

Figure 7.5 Evidence provided by the probability distribution of *Dendrocoelum lacteum* when absent. The distribution has low-information content.

If this method is used, the absent evidence is said to be confirming, since its focus is a singleton hypothesis. The alternative approach is to represent the absent evidence as disconfirming a singleton hypothesis, equivalent to confirming its complement. Rather than use the probability distribution for the absent state, using the distribution corresponding to the complement of this state seemed appropriate, i.e. present, to provide the mass assignment. Since the distribution is usually unimodal the choice of the mass value is clear. The maximum value of the present-state distribution is used to assign a basic probability number to the complement of its associated focus. Referring to **Figure 7.6** which shows the present-state probability distribution for *Dendrocoelum lacteum*, the probability mass m =

0.74 is assigned to $\neg\{B2\}$, i.e. $\{B1a, B1b, B3, B4\}$. Thus the absence of this taxon provides evidence that disconfirms $\{B2\}$ to this degree.

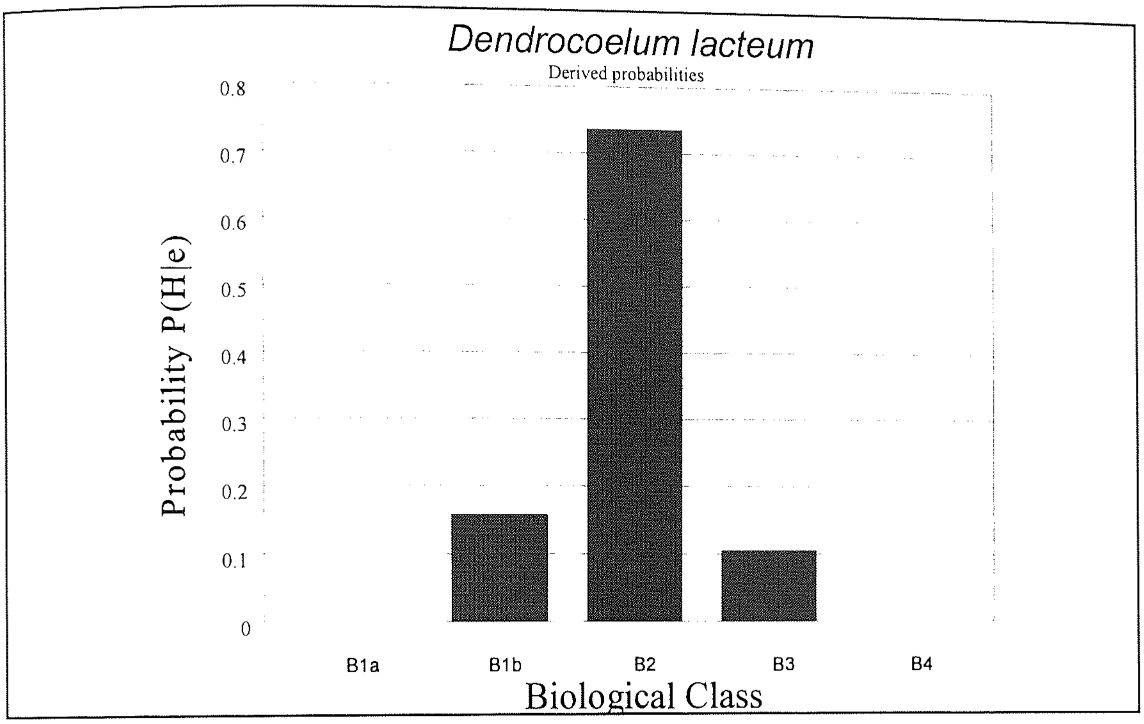

Figure 7.6 Distribution for *Dendrocoelum lacteum* in the present state.

7.2.5.2 Present and Confirming Absent Evidence

The Yorkshire data samples were classified by a Dempster-Shafer system in which the evidence was represented as simple support, with absence confirming a singleton hypothesis. **Table 7.3** shows the confusion matrix obtained from this test. The classification performance for the highest and lowest quality water (B1a and B4) is good, but B1b to B3 sites are consistently misclassified, with a tendency to *upgrade* from the expert's opinion for B1b and B2 sites, a highly undesirable property. A severe misclassification occurred for one B3 site, with the system classifier grading the site at B1a. Over the entire data-set the classification rate was 47.17%, with 28 sites misclassified.

Table 7.3 Classification of Yorkshire data, with evidence represented as simple support. Present and *confirming* Absent data were considered.

		Expert Decision				
		B1a	B1b	B2	B3	B4
Classifier Output	B1a	6	8	5	1	0
	B1b	0	0	0	0	0
	B2	0	2	3	5	0
	B3	0	0	1	8	0
	B4	0	0	1	7	8

Note: Classification rate is 47.17%.

The poor performance arises from the unsatisfactory method in which the singleton focus is chosen to represent the simple support assignment. In spite of the low information content of the absent data, the cumulative effect is to distort the classification, usually towards the extremes of the quality scale. This bias works to the advantage of sites classified by the expert as those classes. For B1a sites for example, the absence of those indicator taxa normally associated with poorer-quality waters reinforces support for B1a, although marginally. The converse is true for B4 sites. Those taxa occurring in the middle range have probability profiles such that their absence provides evidence that simultaneously supports {B1a} and {B4}. In simple support, only one of these foci will receive the basic probability assignment. Given the small differentiation between any of the singletons for the absent state, the choice may seem arbitrary.

The example below (**Table 7.4**) shows the simple support classification of Yorkshire Water site 621A in which absent evidence is used as confirming evidence. The support for each singleton hypothesis is large, leading to a high-level of conflict (log(K.II K_i) = 3.94). Since the evidence is all confirming, the support against each hypothesis c_i is zero. The accumulation of contradictory evidence leads to a majority decision for B1a which however is indefinite (I-value = 9) but is within one class of the expert's opinion (B1b).

Table 7.4 Simple support classification of site 621A using Present and Confirming Absent evidence

LUMBRICULIDAE	ABUNDANT	*Asellus aquaticus*	ESTABLISHED
Gammarus pulex	ABUNDANT	*Baetis rhodani*	ESTABLISHED
ELMINTHIDAE	ABUNDANT	*Ephemerella ignita*	ESTABLISHED
Lymnaea peregra	ESTABLISHED	*Leuctra* spp.	ESTABLISHED
Ancylus fluviatilis	ESTABLISHED	DYTISCIDAE	ESTABLISHED
Sphaerium spp.	ESTABLISHED	*Glossosoma* spp.	ESTABLISHED

(a) Indicator taxa present at the site. (Rare evidence ignored).

EVIDENCE			
Class	For (f)	Against (a)	Conflict (K_i)
B1a	0.977783	0	1
B1b	0.928966	0	1
B2	0.933283	0	1
B3	0.933283	0	1
B4	0.929371	0	1

(b) Evidence for, against, and degree of conflict for each hypothesis (class)

SUPPORT			
Class	For (p)	Against (c)	Residue (r)
B1a	0.977783	0	0.022217
B1b	0.928966	0	0.071034
B2	0.933283	0	0.066717
B3	0.832192	0	0.167808
B4	0.929371	0	0.070629

(c) Basic probability numbers arising from combining evidence in (b).

INTERNAL/OVERALL CONFLICT				
σ	K	$K.\Pi\, K_i$	$\log K$	$\log (K.\,\Pi\, K_i)$
90.1952	8884.59	8884.59	3.9486	3.9486

(d) Conflict values and normalisation factors

BELIEF REPRESENTATION				
Class	Belief	Plausibility	Doubt	Uncertainty
B1a	0.487954	0.499041	0.500959	0.011087
B1b	0.144993	0.156080	0.843920	0.011087
B2	0.155094	0.166181	0.843920	0.011087
B3	0.054983	0.066070	0.933930	0.011087
B4	0.145889	0.156976	0.843024	0.011087

(e) Resulting belief function parameters after combination of all present evidence

Overall conflict levels are very high resulting from (a) the number of sets of evidence being considered and (b) the inherent contradictory nature of the evidence itself. In contrast to singleton support, where the inclusion of absent evidence improves the classification rate, absent evidence represented as confirming simple support leads to serious misclassifications. The mechanism for choosing the focus of support arising from absent evidence may be improved, but eliminating the arbitrariness of this process is difficult, apart from ignoring the evidence altogether as 'unsafe'. Including absent evidence represented as confirming simple support leads to poorer classification performance than present evidence alone, and thus this representation is of little value in this context.

7.2.5.3 Present and Disconfirming Absent Evidence

This part of the investigation was to test the hypothesis that the use of disconfirming, rather than confirming absent evidence should improve classification performance. For disconfirming absent evidence the focus is the complement of a singleton set, rather than the singleton itself. Barnett's scheme can also be used for this evidence representation. A brief example may clarify the idea of disconfirming evidence. Rather than using evidence provided by the absence of a sensor to support a hypothesis directly, we say that it refutes it. Thus the absence of *Sphaerium spp.* refutes (say) {B2}. This can be expressed as a basic probability assignment $m(\neg\{B2\}) = s$, where s is the degree of support against {B2}, or conversely, for $\neg$\{B2\}. Using real numbers, a valid assignment under simple support may be $m_1(\neg\{B2\}) = 0.3$, $m_1(\Theta) = 0.7$, depending on the actual probability profile. The sum of these is unity, as required by the definition of a *bpa*. If now the next sensor data is the abundance of *Leuctra spp.*, a suitable basic probability assignment may be $m_2(\{B1a\}) = 0.6$, $m_2(\Theta) = 0.4$. The orthogonal sum of these two assignments results in *bpa* whose foci are $\neg$\{B2\}, {B1a} and Θ.

Pair-wise combinations of disconfirming probability assignments can be expensive for processing and storage due to the need to enumerate all the subsets and super-sets of a given set. For instance the combination of an assignment to $\neg$\{B1a\} and one to $\neg$\{B1b\} requires the evaluation of the set intersection, in this case {B2,B3,B4}.[2] Barnett's scheme allows the combination of confirming and disconfirming evidence in linear time for simple support functions.

---

[2] Such a procedure was in fact required for the implementation of consonant belief functions.

7.2.5.4 Procedure: Disconfirming Absent Evidence

Samples in the Yorkshire Water data set were classified by a Dempster-Shafer system in which the evidence was represented as simple support, with absent evidence disconfirming. Evidence was combined using Barnett's scheme in sensor-state order, using the original probability distributions as elicited from the domain expert. Even with the use of Barnett's scheme, the time taken to classify a particular sample when absent evidence is considered is appreciable, and is unsuitable for computation in interactive mode.

7.2.5.5 Results

Contrary to expectations the inclusion of *disconfirming* absent evidence resulted in a reduction in the classification rate compared with confirming absent evidence (**Table 7.5**). The confusion matrix shows misclassification for all except B4 sites, with three B1a sites were seriously misclassified as B4.

Table 7.5 Classification of Yorkshire data with evidence represented as simple support. Present and *disconfirming* Absent data were considered

		Expert Decision				
		B1a	B1b	B2	B3	B4
Classifier Output	B1a	1	0	0	0	0
	B1b	2	1	1	0	0
	B2	0	2	2	1	0
	B3	0	0	3	7	0
	B4	3	7	4	13	8

Classification rate 35.85%.

A consequence of representing evidence as disconfirming is the accumulation of evidence against hypotheses in the frame of discernment. This contrasts with all-confirming evidence in which support accrues only *for* each singleton. Examination of intermediate calculations for the three B1a sites misclassified as B4 revealed high degrees of internal conflict resulting from evidence both for and against the same singleton hypothesis.

Table 7.6 shows intermediate calculations for one of these sites (553B). Support both for and against each singleton produces an internal conflict factor K_i associated with each singleton. The high conflict factor for class B1a results from the fact that there is a large amount of evidence that both supports and refutes it. Support p for each hypothesis is then calculated by combining the $n = 5$ basic probability assignments represented by f and a.

Table 7.6 Simple support classification of Yorkshire site 553B. Absent evidence disconfirms

EVIDENCE			
Class	For (f)	Against (a)	Conflict (K_i)
B1a	.954839	.990819	18.543260
B1b	.803580	.985923	4.813914
B2	.555556	.999341	2.248150
B3	.000000	.934031	1
B4	.000000	.581818	1

(a) Evidence for, against, and degree of conflict for each hypothesis (class)

SUPPORT			
Class	For (p)	Against (c)	Residue (r)
B1a	.162562	.829749	.007689
B1b	.054453	.932237	.013310
B2	.000822	.998520	.000658
B3	.000000	.934031	.065969
B4	.000000	.581818	.418182

(b) Basic probability numbers arising from combining evidence in (a).

The overall belief parameters for this example are such that the evidential width is not identical for each singleton hypothesis. Combination of disconfirming evidence leads to belief residing with subsets of Θ which are also supersets of the singletons, i.e. the belief associated with the complements of the singletons.

7.2.5.6 Use of variable data-quality

The interpretation of benthic data arising from the absence of an indicator is problematic, in that in this representation the absence of taxa associated with higher-quality water has equal weighting with those of lower-quality. However, since higher-quality taxa are more likely to be absent from riffles due to factors other than pollution, it would seem reasonable to weight the evidence from their absence less than that of lower-quality taxa. These ideas were tested via the use of a variable data-quality weighting shown in **Table 7.7**.

Table 7.7 Variable data-quality weighting for absent taxa

Simple support focus	B1a	B1b	B2	B3	B4
Data quality of absent evidence	Poor	Fair	Good	High	Certain

This weighting was applied first of all to the B1a sites, five of which were previously seriously misclassified, and then to the B3 sites to ensure that this weighting had not resulted in a worsening of classification performance. As before, absent evidence was represented as disconfirming simple support.

Table 7.8 shows that simple support classification for B1a sites is improved, being comparable with the situation in which present evidence alone is considered. This supports the expectation that the absence of higher-quality taxa should be weighted less than poorer-quality taxa. For the B3 sites, misclassifications to B2 occurs for the same reasons observed for singleton support: those sites classified by the expert as B3 contained taxa which gave strong support to B2. This may be interpreted as an inconsistency in expert diagnosis, or, more likely, the failure of the classifier to capture data weighting induced by certain taxa occurring together.

The classification of site 547B, a B3 site in the expert's opinion, exhibits an interesting phenomenon which occurs due to the combination of evidence which both confirms and disconfirms singleton hypotheses. Evidence for B3 ($f = 0.76$) is more than counterbalanced by evidence against it ($a = 0.92$). No evidence either accrues for or against B4, and yet this receives the overall majority support. This arises from the uncommitted belief which remains with the environment Θ, coupled with the evidence *against* all the other hypotheses, resulting in a basic probability mass assignment remaining with B4 (since $\{B4 \cap \Theta\} = B4$).

Table 7.8 Effect of applying variable weighting to disconfirming absent evidence

		Expert Decision			
		B1a		B3	
		Zero weight	Variable weight	Zero weight	Variable weight
Classifier Output	B1a	1	4	0	0
	B1b	2	2	0	0
	B2	0	0	1	4
	B3	0	0	7	7
	B4	3	0	13	9

Notes: The test was carried out on Yorkshire Water sites classified as B1a and B3 by the domain expert. Evidence induced by the absence of taxa associated with higher-quality waters received less weight than those associated with lower-quality. The weighting was progressively increased from B1a to B4.

7.2.6 Alternatives to Dempster's Rule

7.2.6.1 Theoretical Background

As discussed in Chapter 3 Dempster's rule has been criticised for giving counterintuitive results after normalisation for highly conflicting evidence (Zadeh, 1986). Because of this phenomenon various workers have considered alternatives or modifications to Dempster's rule, or examined its theoretical justification (Voorbraak, 1991). Gordon and Shortliffe (1985) suggest that in combining conflicting evidence the mass of the null set could be assigned to Θ, rather than used to normalise the focal masses. Detailed studies of belief combination using conflicting evidence have been carried out leading to the development of 'ideal' combination rules (Cheng and Kashyap, 1988; Chang and Kashyap, 1990). For evidence expressed in interval form, such as [0.15,0.25] or [0.81,0.9] the rules should not only satisfy properties such as closure, commutativity, associativity, but also maintain information on the conflicting nature of the evidence.

Thus, if two sets of evidence are in conflict, the resulting uncertainty in the combined evidence should be greater than that expressed in the original intervals. This is not so with Dempster's rule, in which the belief interval $Pls(A)$ - $Bel(A)$ for some proposition A is progressively reduced as evidence is combined whether or not the evidence before combination was in conflict. Conversely, non-conflicting evidence should reinforce, leading to lower uncertainty (and therefore belief intervals) in the combined result. Chang and Kashyap (1990) have developed a rule (known as the T-R combination rule) which they claim satisfies the basic axioms and the additional conflict-resolution and reinforcement properties.

The T-R combination rule is developed using geometric arguments by first considering the interval [a,b] as a vector in a two-dimensional coordinate system, so that all evidence can be represented within a triangular region (**Figure 7.7**). A *discrimination measure* is defined to be $a + b - 1$, the support for the hypothesis H minus the support for $\neg H$, the negation of H. If two sets of evidence have discrimination measures of opposite signs, they are considered to be in conflict. Geometrically, [a,b] and [c,d] are in conflict if vectors (a,b) and (c,d) do not fall in the same triangle BCD and ACD in **Figure 7.7**.

Evidential intervals are mapped to a rectangular region in such a way that one component of each interval in this region is equal to the discrimination measure, thereby incorporating a means of detecting and resolving conflict.[3] The rectangular maps are then combined using functions which preserve associativity and commutativity. Finally, the resulting rectangular co-ordinates are mapped back into the original co-ordinate system to

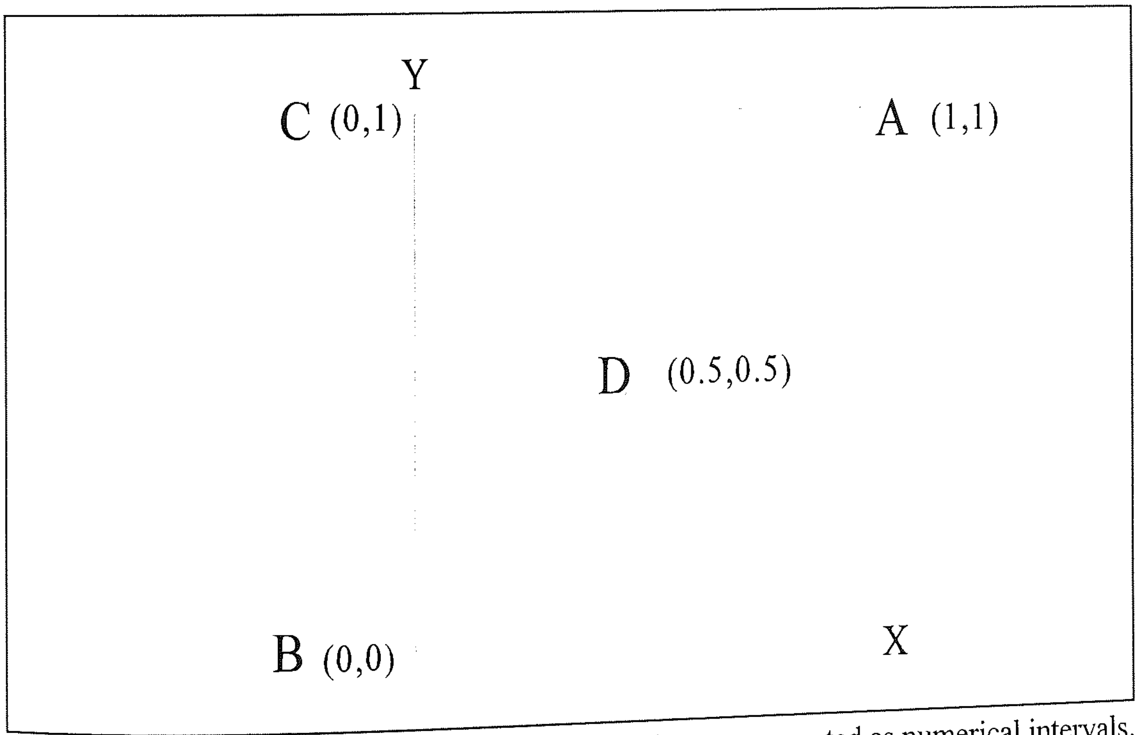

Figure 7.7 Triangular region used to represent evidence represented as numerical intervals. The intervals are vectors bounded by the region ABC.

[3] Mathematically, this procedure is known as a homeomorphic transform.

yield the combined evidential interval. Details of the actual rule are given in an appendix.

Consider for example a situation in which support for an hypothesis from one piece of evidence is expressed in interval form as [0.2,0.4], while according to a second set of evidence the support is [0.7,0.9]. The two sets of evidence are clearly highly conflicting. Combination of these using Dempster's rule leads to the evidential interval [0.57,0.64], a decisive result within the range [0.5,1.0] with a width of 0.07, considerably less than those of the original components. Chang and Kashyap claim that this behaviour is unacceptable. Their new combination rule would yield an interval of [0.48,0.74]. The T-R rule has the reinforcing properties of Dempster's rule, but also a property of increasing evidential width for conflicting components.

7.2.6.2 Procedure

The simple support representation appears to lend itself to Chang and Kashyap's combination scheme for interval-based evidence, since support of degree s for a singleton proposition can be expressed in interval form as $[s,1]$. A Dempster-Shafer classifier was developed which used simple support functions combined using the T-R combination rule rather than Dempster's rule as incorporated in Barnett' scheme. The 53 sites of the Yorkshire Water data were classified using present (established and abundant) evidence under these conditions.

7.2.6.3 Results

One problem that arises from using this rule is determining the manner in which it is used in a multi-hypothesis space, and how the evidential intervals can be used to reach a decision. For the Bayesian and Dempster-Shafer calculi, the probabilities or mass assignments are normalised to sum to unity. Since the T-R rule has been designed to avoid normalisation, the evidential intervals for each singleton are unnormalised, as the example given in **Table 7.9** shows.

Table 7.9 Example of resulting belief function parameters after combination of evidence using T-R combination rule

BELIEF REPRESENTATION				
Class	Belief	Plausibility	Doubt	Uncertainty
B1a	.997514	1.000000	.000000	.002486
B1b	.999957	1.000000	.000000	.000043
B2	.367630	1.000000	.000000	.632370
B3	.367630	1.000000	.000000	.603550
B4	.000000	1.000000	.000000	1.000000

This suggests that the evidence simultaneously gives highly definite support for both B1a ([0.9975,1]) and B1b ([0.999,1]). In this case the decision could be given as B1b+ or perhaps B1b/B1a for the particular site. To assess the classification performance over the entire data-set, a "highest-wins" decision mechanism was adopted to select the proposition with the maximum support, as before.

The classification performance over the data-set is shown in **Table 7.10**. The decision mechanism forces automatic alignment on a class, corresponding to the criterion for evaluating the confusion matrix. The error rate is 47.16%, comparable to simple support in which present and confirming absent data were combined.

Table 7.10 Classification of Yorkshire data under T-R combination rule

		Expert Decision				
		B1a	B1b	B2	B3	B4
Classifier Output	B1a	5	6	3	0	0
	B1b	1	1	0	0	0
	B2	0	1	6	10	0
	B3	0	0	1	9	4
	B4	0	0	0	2	4

Notes: Only evidence from taxa present in each sample was considered. Classification rate 47.16%.

In fact the performance on a sample-by-sample basis is better than that suggested by the confusion matrix alone. **Table 7.11** shows details of the classification decisions for three Yorkshire sites, which count as misclassifications for purposes of computing the confusion matrix. For sites 495B and 640, the classifier shows close agreement with the expert's classification of B3+ and B1b respectively. However the decision parameters for site 443C indicate why the T-R combination rule can be problematic for a decision maker: classes B1a to B3 receive comparable degrees of support.

Table 7.11 Decision parameters for three sites considered as misclassified under T-R combination

	Di	Decision Order	Di	Decision Order	Di	Decision Order
B1a	0.992822	1	0.356825	3	0.998659	1
B1b	0.970767	2	0.317105	4	0.971169	3
B2	0.894427	3	0.893499	1	0.957605	4
B3	0	4	0.886555	2	0.994319	2
B4	0	4	0	5	0	5
Expert's Classification	B1b		B3+		B2+	
Expert Base Class	B1b		B3		B2	
System Base Class	B1a		B2		B1a	
Nearest Rank	B1a		B2		B1a	
SCI	0		2		0	
Indicator Value	19		17		19	
Site Reference	640		495B		443C	

7.2.6.4 Discussion

The behaviour of the T-R rule depends very much on the initial interval. For the simple support functions, the upper value of the interval is 1, corresponding to the plausibility of each singleton proposition. (This follows from the fact that, for a particular piece of evidence, only one singleton or its negation receives support). In this case the plausibility will always be unity after each T-R combination. For instance the combination of the intervals [0.7,1.0] and [0.3,1.0] results in [0.9012,1.0]. If however the second component is represented as [0.3,0.5], the resulting interval is [0.796,0.877]. The very different result arises from the use of a discrimination measure $a + b - 1$ in the T-R rule derived from considering which half of the interval [0,1] the interval [a,b] lies. Since [0.3,0.5] lies in the interval [0,0.5], its discrimination measure is negative, while [0.3,1.0] has a positive measure.

In the case of simple support derived from the sensor probability distributions, all the discrimination measures for the intervals will be positive. This leads to over-reinforcement of evidence using T-R rule combination. Using Dempster's rule, the combination of [0.7,1.0] and [0.3,1.0] results in [0.79,1.0], reflecting the slight increase in information regarding the truth of the proposition afforded by the weak second component. This contrasts strongly with the T-R rule which produces a highly definite result approaching absolute certainty. The gradual reinforcement property of Dempster's rule is desirable and intuitive. For simple support, the form of the evidential intervals derives from the need to maintain a basic probability assignment. It is possible that this constraint need not be observed for the T-R combination rule, so that a domain expert may be free to express belief in a proposition using any sub-interval of [0,1].

The use of the T-R rule is problematical in a multi-hypothesis space since it reflects internal conflict, i.e. between evidence supporting and refuting the same hypothesis. It is not clear how this extends to conflict between the various propositions. For a large number of sensor data, much of which will contain conflicting evidence by supporting exclusive propositions, Dempster's rule maintains a more consistent probability assignment via the mechanism of normalisation.

7.2.7 Discussion: Simple Support

The simple support representation of belief in the Dempster-Shafer theory in which belief is focused on singletons or their complements allows the use of Barnett's scheme for rapid evidence combination. This is particularly beneficial when combining disconfirming evidence. For classifications on the Yorkshire Water sites the use of only present data gave a classification rate of around 55%. The addition of absent evidence, resulted in large misclassifications, particularly when used as confirming evidence. This contrasts with the Bayesian or Dempster-Shafer classifiers using singleton support, in which classification performance over the data-set declined as absent evidence was increasingly discounted.

This suggests that a better biological classification performance can be obtained by the considering evidence from only those indicator taxa present at a site. Apart from the computational advantages, the simple support representation embodies evidential discount naturally. Unless the entire probability mass can be focused on a singleton hypothesis, in which case the decision is unanimously for that proposition, there will always be a degree of uncommitted belief assigned to the environment Θ. This ensures that conflict can be managed successfully. The addition of absent evidence decreases the role of Θ.

The simple support representation is also attractive from the point of view of knowledge acquisition of the probability measures. For the Bayesian or singleton support belief functions, the expert is obliged to supply a discrete distribution across the five biological classes, a task that can be onerous. For simple support, a domain expert could be asked to indicate the one quality class with which the benthic sensor state is strongly associated, and to what degree. This would reduce to stating two probability measures for the states of established and abundant. This representation is more likely to be successful for those sensors with pronounced indicator values, i.e. those definitely associated with one or two adjacent classes. Alternatively, a domain expert could consider the presence of certain taxa to indicate disconfirming evidence, if those taxa were considered weak or ubiquitous indicators. In this scenario, the role of disconfirming evidence would arise without the need to consider absent taxa, which for the simple support representation distorts the evidence provided by those present.

The investigation of simple support looked at a particular alternative to Dempster's rule of combination, the so-called T-R combination rule, that appeared at the outset to be a viable means of combining simple support functions. However because of the nature of the belief intervals constructed from the basic probability assignments, the combination procedure over-reinforces evidence. This, coupled with the absence of normalisation to maintain consistent probability mass across the propositions, makes it difficult to decide the outcome of the classification using this technique. The performance was judged by choosing the class with the highest belief as representative of the classification. On this basis, the classification rate was comparable with simple support in which present and confirming absent data was included. For classification of an individual site, distinguishing between support for competing hypotheses may be difficult for a decision-maker.

7.3 Classification Using Consonant Belief

7.3.1 Theory

Bayesian and singleton support functions give simultaneous support to more than one singleton hypothesis. Although singleton support functions can be decomposed into simple support functions, this is not possible with Bayesian belief. Intuitively, this follows from the dissonant nature of the Bayesian distribution. No uncertainty is present in the assignment: all the probability mass is focused on one or more of the singleton hypotheses. According to Caselton *et al.* (1988), since these mutually exclusive states are being *simultaneously* supported to some degree the Bayesian belief function can be viewed as inherently contradictory, and the quality of evidence in this form may therefore be suspect.

More confidence may however be associated with evidence that supports several conclusions in general agreement with each other. Consider the following *bpa* graphically depicted as shown in **Figure 7.8**. The diagram represents the *bpa* $m(\{B1a,B1b,B2\}) = 0.7$, $m(\{B1b,B2\}) = 0.2$, $m(\{B2\}) = 0.1$. The least precise proposition {B1a,B1b,B2} receives the greatest credence, which diminishes as the propositions become more specific. Intuitively, the more precise the proposition, the less certain we are about its credibility. Belief functions that can be structured this way are said to be consonant or resonant (Dong and Wong, 1986a, 1986b).

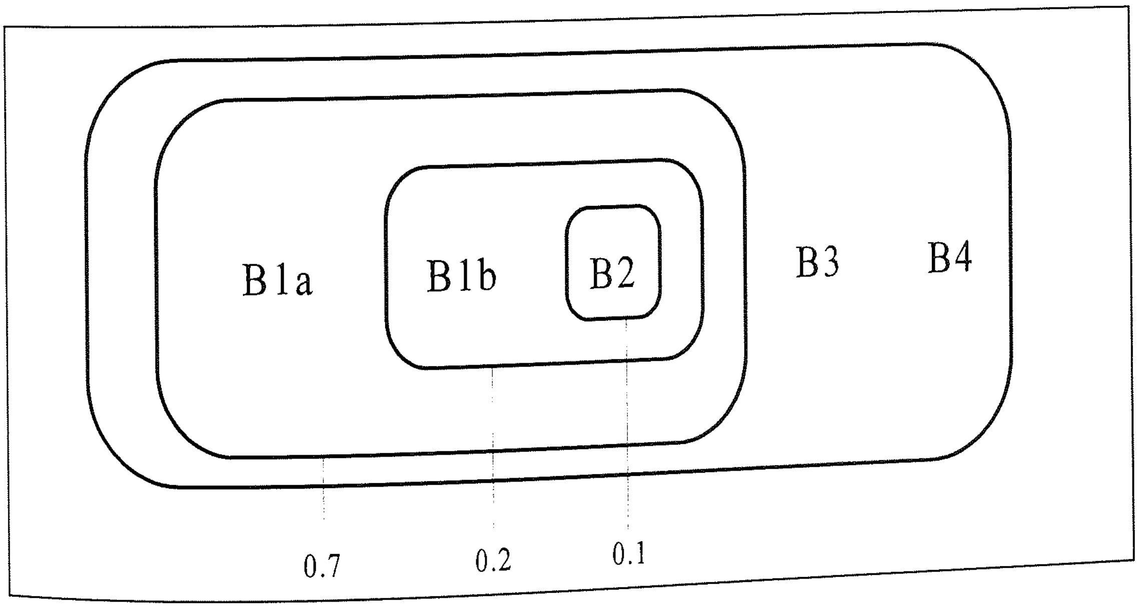

Figure 7.8 Consonant belief function

Formally, if A_1, A_2, A_3 are elements within the frame of discernment, a consonant belief function is so ordered that

$$Bel(\{A_1\}) < Bel(\{A_1,A_2\}) < \ldots < Bel(\{A_1,A_2,\ldots,A_n\}) \qquad \textbf{(7.5)}$$

In the domain of biological classification the observation of the state of one sensor provides a set of discrete probabilities across each of the singleton hypotheses in the frame of discernment. Thus, if e is the sensor state and H_i is the water quality class, the observation of that state provides the discrete probability distribution $p(e|H_i)$ for $i = 1 \ldots n$. Adopting a consonant belief structure for these sample likelihoods allows one to compute the belief functions for each of the nested subsets (Shafer, 1976):

$$Bel(A) = 1 - \frac{\max\limits_{H_i \in \neg A} p(e|H_i)}{\max\limits_{H_i \in \theta} p(e|H_i)} \qquad \textbf{(7.6)}$$

Caselton *et al.* (1988) claim that belief functions of this form are appealing for engineering applications.

Consonant belief functions are closely allied to studies of hierarchical diagnostic spaces (i.e. those in which hypotheses are related hierarchically). The Dempster-Shafer scheme is well-suited to modelling diagnostic reasoning in such domains: indeed Gordon and Shortliffe (1985)

"... are unaware of another model that suggests how evidence concerning hierarchically-related hypotheses might be combined coherently and consistently to allow inexact reasoning at whatever level of abstraction is appropriate for the evidence that has been gathered."

Since Barnett's scheme is available for only singletons and their negations, these same authors have extended the scheme to develop an approximate method for managing hierarchical reasoning (Gordon and Shortliffe, 1985). Since then Shafer and Logan (1987) have developed an exact algorithm for the implementation of Dempster's rule for such evidence, while Hau (1990) shows that belief functions associated with a hierarchical hypothesis space are separable. Even in applications where the hypotheses are not naturally hierarchically-ordered (such as the biological water classes), evidence at different levels of abstraction can occur. The decomposition algorithms of Hau and the computational schemes of Barnett (1981) and Gordon and Shortliffe also provide a means of retracting evidence from faulty or suspect sensor data.

7.3.2 Procedure

To investigate these ideas the classification of the Yorkshire Water data-set was carried out in which the sensor evidence was represented as consonant belief. Physically, each member of the power set for the frame of discernment is represented by its characteristic function. For example the characteristic function for {B1a,B1b,B2,B3,B4} is {1,1,1,1,1} while that for {B2,B3} is {0,0,1,1,0}. The 1 or 0 therefore denotes whether or not a singleton hypothesis within the environment Θ is represented by the particular set. Using this binary scheme, there are 2^5 characteristic functions in the power set. This representation allows the set intersections arising from the application of Dempster's rule to be rapidly determined by bitwise-and operations between the characteristic functions. For example, consider a mass assignment to the set {B1a,B1b,B3} and a second to {B1b,B3,B4}. The orthogonal sum of these assignments will be made to the intersection of these two foci. Representing these by their characteristic functions, the intersection is {1,1,0,1,0} & {0,1,0,1,1} = {0,1,0,1,0}, i.e. {B1b,B3}.

For any particular piece of sensor evidence, the consonant sets and the belief assignment are derived by first assigning the entire belief to the superset that spans the range of the probability distribution. Evidence with a pronounced indicator value may have the distribution focused on two or three adjacent classes. A sensor in the abundant state may have a distribution spanning {B1b,B2,B3}, for which the belief assignment *Bel*({B1b,B2,B3}) = 1 follows. The first consonant subset below this superset is found by eliminating the singleton set with the lowest probability value within the current range. This process continues until the lowest level is reached, normally at the singleton set situated at the modal value. At each level, belief is assigned according to equation **(7-6)**.

This scheme also provides a more natural representation for distributions that are not unimodal, such as those for which evidence suggested by the absent sensors or for ubiquitous taxa whose probability values are equal across a range of adjacent classes. If no singleton focus can be discerned, belief is assigned to that superset whose associated probability values are equal.

LEONARDO's list-processing capabilities allow a straightforward implementation of operations between sets representing incoming sensor evidence and existing target sets (i.e. those focal elements for which a probability mass is currently assigned). Duplicate list elements are automatically eliminated between set operations, leaving only those sets that are eligible for subsequent operations as future targets. Normalisation involves removing the null set from the list of target sets and reassigning its mass to the proper focal elements. For the decision the plausibilities of each singleton hypothesis H_i are obtained by calculating $1 - Bel(\neg\{H_i\})$, requiring the summation of the mass assignments associated with subsets that intersect H_i.

The classification was carried out for (a) present data and (b) present and absent data. Indicator taxa whose state was rare was considered to have neutral effect on the decision. One consequence of the need to enumerate all the supersets that arise from consonant belief functions which span the frame (as with absent data) is the length of time to carry out the evidential combinations. The average time to classify a sample using only present data was in the region of 12 to 15 seconds. This increased dramatically when absent evidence was also considered, approaching 10 hours[4] to classify the 53 samples in the set. Clearly this is not suitable for interactive computing, and the classification was carried out in batch mode.

7.3.3 Results

Table 7.12(a) shows the classification errors obtained for the Yorkshire Water data set considering present data only, using consonant belief representation derived from the area-adjusted distributions. The classification rate of 58.59% compares with the singleton support classification of 60.38% under the same conditions. When absent evidence is also considered, the classification rate improves considerably (**Table 7.12**(b)), and is identical to the singleton-support classification using area-adjusted distributions.

---

[4] These figures refer to computation times on a 4DX-33MHz personal computer.

Table 7.12 Classification of Yorkshire data with evidence represented as consonant belief

Classifier Output		Expert Decision				
		B1a	B1b	B2	B3	B4
	B1a	2	2	0	0	0
	B1b	4	4	1	0	0
	B2	0	2	9	7	0
	B3	0	0	0	14	6
	B4	0	0	0	0	2

(a) Evidence from taxa present (only) in each sample was considered. Classification rate 58.59%.

Classifier Output		Expert Decision				
		B1a	B1b	B2	B3	B4
	B1a	4	2	0	0	0
	B1b	2	4	1	0	0
	B2	0	2	8	1	0
	B3	0	0	1	17	0
	B4	0	0	0	3	8

(b) Evidence from taxa present in and absent from each sample was considered. Classification rate 77.36%.

7.3.4 Discussion

With the inclusion of data from taxa absent from the sample, the classification performance is markedly improved, but at considerable computational effort. This results from the need to enumerate the hierarchical set of subsets for each piece of evidence and to compute the orthogonal sum arising from their interaction with the current focal elements. One problem with this type of belief representation is that consonance is not preserved after evidential combination. Consider the combination of evidence for a particular site 443C in the Yorkshire Water data set. For the first piece of evidence, the observation that *Potamopyrgus jenkinsi* is established at the site, a consonant set {{B1a,B1b,B2,B3},{B1a,B2,B3},{B1a,B2},{B1a}} can be constructed from the probability distribution derived from the elicited histograms. The use of the area-adjusted distributions also results in Θ being part of the consonant set due to the redistribution of probability mass across the propositions. Thus, this adjustment exacerbates the problem of subset enumeration.

Combining consonant evidence from *Bithynia tentaculata*, *Lymnaea peregra* and *Ancylus fluviatilis* results in mass assignments to 17 subsets of Θ, including {B1b}, {B2}, {B3}, {B1b,B2}, {B1b,B3}, {B2,B3} and {B1b,B2,B3}. Further evidence from the presence of *Tubificidae*, *Helobdella stagnalis* and *Erpobdella octoculata* eliminate support for {B1b} and its supersets. Evidence begins to accumulate towards the lower-quality classes {B2,B3,B4} and {B3}. Sensor data from *Hydracarina*, an indicator of good-quality waters, provides consonant support for {B1a,B1b,B2} and {B1a,B2}. However the combination of this current support with the target sets results in support only for {B2}, all other evidence accumulating in the null set. The evidence from this last sensor for the higher quality {B1b} has no effect since this proposition has been eliminated. Since the mass of the null set will usually be quite large after many set intersections, normalisation of belief results in considerable support for the remaining focus {B2}. In this example, subsequent evidence supporting multiple hypotheses serves only to promote {B2}. The result is a highly-focused decision for this class, with an *I*-value of 19.

This phenomenon of promoting the "middle ground" was the basis of Zadeh's objection to Dempster-Shafer normalisation, and is a common occurrence in the combination of consonant belief functions. Area-adjustment of the distributions can exacerbate the problem by promoting those propositions occurring at the intersection of conflicting evidence. These observations lend support to the idea that the device of evidential discounts should be used here as in singleton support, to ensure that Θ prevents the premature elimination of competing propositions.

7.4 Summary

This chapter has described in detail the experimental programme that investigated two forms of belief function: simple support and consonant belief. These are distinctly non-Bayesian and allow greater flexibility in the way belief is represented and assigned for uncertain reasoning. The Dempster-Shafer calculus provides a coherent combination scheme for integrating evidence at different levels of abstraction within the domain of interest.

Simple support functions represent the most direct form of basic probability assignment. One difficulty that persists with this method when used for biological classification is the choice of the focal element to receive support, and is particularly problematic when used to represent evidence from absent taxa. Two variants of assignment were used for absent evidence: the first to confirm support for a singleton focus, the second to disconfirm, equivalent to confirming its complement. In contrast to its use when represented as singleton support or Bayesian belief, the inclusion of absent evidence resulted in a degradation of classification rates compared with the consideration of solely present data.

Contrary to expectation, the use of disconfirming evidence did not enhance classification rates compared with the inclusion of confirming absent evidence. The application of variable weighting, in which the absence of higher-quality taxa was weighted less those associated with poorer-quality waters did however reduce the serious misclassifications that occurred without weighting.

The investigation of simple support concluded with the use of an alternative to Dempster's rule of combination, the T-R rule due to Chang and Kashyap (1990). The rule's avoidance of normalisation can lead to large degrees of support simultaneously for rival hypotheses, which renders the decision process problematic.

Consonant belief functions are intuitively attractive for representing evidence at levels of abstraction other than singletons or their complements. In computing terms however, they introduce a level of complexity for which the Dempster-Shafer calculus has been criticised. The results suggest that, when considered over the Yorkshire Water sites, the classification performance compares well with the Bayesian or Dempster-Shafer classifiers using singleton support, but the considerable extra computational effort required for consonant belief combination is not justified.

Chapter 8

Recommendations for Further Work

8.1 Introduction

This chapter discusses suggestions for further work on biological classification using uncertain reasoning methods. Within this context a design of a decision support system that could incorporate automated biological classification and heuristic or model-based knowledge on biological surveillance is presented. The possibility of combining expert opinion on probabilistic knowledge of indicator taxa is discussed, and concludes by identifying Bayesian networks for developing more complex probabilistic reasoning systems for this domain.

8.2 Decision Support System for River Water Quality

8.2.1 Rationale

The starting point for the work described in this thesis was the possibility of building a complete decision support system for assessing river pollution. The facility to produce a classification of biological quality, which has been the focus of this study, would form an important component of such a system. This section presents an overall design of the so called BERT (Benthic Ecology Response Translator) system. It begins with an outline "requirements definition" for the proposed software. Such a definition, for any large software system, provides a summary and reference for the benefit of both users and developers regarding the services to be provided by the software.

8.2.2 Outline Requirements Definition

The BERT system is required to help in the interpretation and analysis of data obtained from biological surveillance at freshwater sites. The system will aid in the identification and location of pollution incidents which may adversely affect water quality. It will be useful in the wider role of assessing and managing changes in river water quality at such sites over time. In broad terms, the requirements of the software system are:-

(i) it is required to give a diagnosis of the river water quality at a sample site in terms of a biological classification system

(ii) it should identify the type of pollution causing a degradation in river water quality at a site

(iii) it will detect trends in the river water quality at a site over time

(iv) it will detect quality trends over a river network and reason on the likely location(s) of incidents of pollution.

8.2.3 Development Environment

Decision-support systems (DSS) typically use a range of software technologies for solving problems in some domain of interest. Examples include those of the OASIS decision-support system for modelling ground-water contamination, integrating ground-water models and chemical and geological databases to form an "expert consultant". A similar paradigm underpins the RAISON expert system, a software package that uses maps, statistics and simulation models to investigate a variety of environmental problems including acid rain and mine effluent (Lam *et al.*, 1989a, 1989b, 1990).

The LEONARDO development environment allows the developer to include causal, algorithmic and frame-based knowledge. It permits execution of external programs, access to spreadsheets and construction of effective user-interfaces. This environment, which was chosen to develop the biological classifiers, would be suitable for the incremental development of an overall decision support system.

8.2.4 Incorporation of heuristics

A combination of rule-based and frame-based knowledge could incorporate heuristics for biological surveillance that could support the automated classification. The automatic classifiers use data from the indicator group only; within a general decision support system, information from the entire sample data could be used to supplement the biological classification to provide "point interpretations". These could give the user qualitative descriptions of overall quality and provide reasons for any degradation.

During conversations with the domain expert rules relating to particular benthic taxa were discovered. The heuristics are formulated here as LEONARDO production rules. For instance:

```
if Plecoptera are Present
then Water_quality is Good
```

or

```
if Chironomus riparius is Abundant
```

then Organic_pollution is Severe

Such rules allow a user to interrogate the knowledge base to find the reasons for reaching a particular qualitative decision, by the process of backward chaining. Within a LEONARDO knowledge base, specific rules such as these could be evaluated within frames for particular benthic invertebrates. Another rule relates to the type of pollution suggested by the complete absence of insects. As a LEONARDO quantification rule, a possible formulation is:

for all Insecta
    If State: of Insecta is Absent
    then Pollution_type includes Insecticide;
    Toxic_pollution is Highly Likely

where the values of the objects 'Pollution_type' and 'Toxic_pollution' are assumed to be sought as sub-goals with the knowledge base. Here the class object 'Insecta' is a member of the superclass Taxon, inheriting its attributes. This could be stored as slots within in a frame (i.e. the attributes of a benthic taxon), and inherited by all members of that class.

Other rules could deal with absent evidence. The absence of taxa may be caused by factors other than water quality, such as predation or even sampling error. Presuming that this data is available during a knowledge-base consultation, the following rule could be used to decide whether the absence of taxa should be taken into account:

If state: of Taxon is Absent
and predation is Unlikely
and sampling_error is Unlikely
and seasonality: of Taxon is In_Season
Then absence: of Taxon is Significant

The group of indicator taxa could be structured in a knowledge base as a subclass of Taxon. If the absence of any of the indicators was significant according to this or similar rule, its evidence could be included in the biological classification decision process.

8.2.5 System services

8.2.5.1 Biological Classification

The system will give a biological classification of the river water quality existing at the site for a particular sample. The user may enter a new sample or choose from an existing sample within the database. The results of the interpretation will be stored. The point interpretation at a sample site uses the direct interpretation, with calculations of biotic indices, to decide the most likely water quality class for the site, and to reason about the most likely form of pollution (if any) present.

8.2.5.2 Spatial and temporal trends

The system will report on trends in river water quality at a site over time, by comparing site point interpretations at different times. Comparison of contemporaneous point interpretations at different sites over a river network will be used, in association with data on sources of pollution, to find potential discharges. A combination of reports on spatial and temporal trends will be used to provide an overall report on river water quality existing over selected sections of the network.

8.2.6 Database design

The data entities within the BERT system, and their attributes, can be identified as follows:

River network N°., Connectivity (i.e. schematic description of the network), Name.
River N°., Name, { Site N°s }, { Pollution-source N°s }, River-network N°.
Pollution source N°, Type [Point/Diffuse], Pollutant-type [Toxic/Organic/Physical], River N°.
Site N°, Name, Map-reference, Position-in-river, Classification [Riffle/Pool], River-zone [Rhithron/Potamon], River N°, { Sample N°s }.
Sample N°, Date-stamp, Site N°, { Benthic list }, Biotic scores { TBI, Diversity }, Biological Classification, Point Interpretation.
Taxon Class N°, Name, (Common name), Taxonomic-level, Probability-data for quality-classes, Biotic-score ratings (TBI, BMWP score), Number at which taxon is Abundant/Rare, Response to Organic/Toxic/Physical pollution.

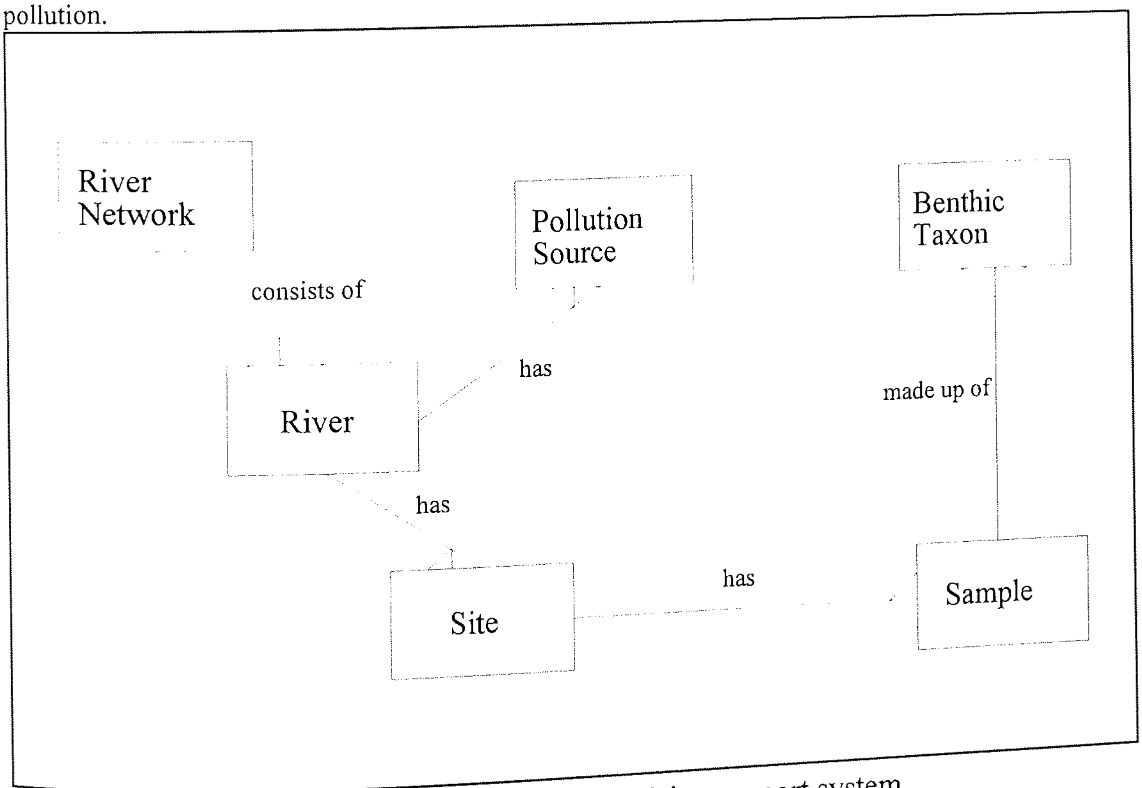

Figure 8.1 Data entities for proposed BERT decision support system

These entities could be constructed as frame objects within a knowledge-based system or

These entities could be constructed as frame objects within a knowledge-based system or stored within an external database. Each of these structures could be expanded to store other attributes of interest. **Figure 8.1** shows an entity-relationship model for the data.

8.2.7 Functional design

Using the data model developed above, a functional design of the software system was developed by the author. A structure chart for this design is given in **Figure 8.2**. The modular nature of the design lends itself to incremental software development.

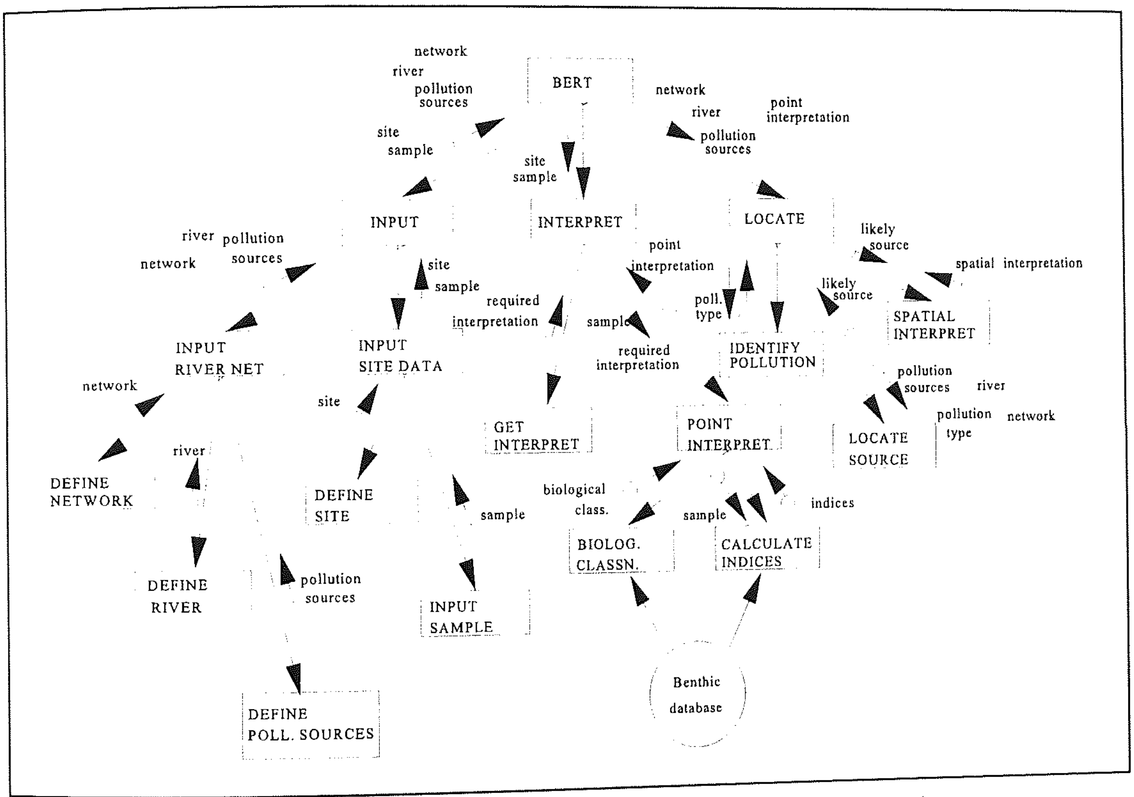

Figure 8.2 Software structure for proposed BERT decision support system

8.3 Expert consensus and knowledge acquisition

Direct elicitation of the discrete probability distributions for the sensor evidence was described in Chapter 4. The probability measures were one expert's opinion of the probabilities of a set of hypotheses H_i given the occurrence of sensor evidence e, for a particular set of benthic indicators. Automated reasoning systems use the evidence provided by this indicator group to arrive at biological classification, whose correctness is judged by its proximity to the expert's own assessment.

The wider acceptance of biological classification using uncertain reasoning methods would require agreement in several areas. For instance, two or more experts in biological surveillance may draw up probability distributions for the same indicator taxon that differ either in profile or degree. These opinions could be pooled by a variety of techniques including the linear or logarithmic pools, or Dempster's rule.[1] Alternatively the use of the simple support belief representation may ease the process of knowledge elicitation by requiring the identification of a single 'preferred' biological class for a particular sensor state. The selection of the indicator taxa could also be optimised for particular regions of the country based on their frequency of occurrence and indicator value. Thus the indicator group for East Anglia would differ say, from one associated with an upland area.

There would also need to be agreement on the sensor states. This may differ from region to region, so that several abundance levels may be identified (e.g. present, few, common, abundant, very abundant) rather than the three levels of presence identified for this project during the knowledge elicitation. One important feature of this elicitation was the realisation that as indicators of river water quality the abundance levels are sensor-specific, so that for instance the term "abundant" for Plecoptera implies different numbers of occurrences than for Tubificidae. Thus, these numbers should be identified for each taxon within the indicator group for the agreed sensor states.

There is also a need to evolve an agreed definition of the biological classes. As discussed in Chapter 2 water companies have in the past used biological classification in which the classes are defined in terms of expected fauna. Observations of the actual fauna are then used to decide the biological class of the sampled site. This however raises the question of to what degree the observed fauna satisfy the quality standard defined by the expected fauna for each class, or how anomalies in presence or absence are accommodated. The approach taken by the RIVPACS project is to use statistical techniques to predict expected fauna for sites of pristine quality based on environmental variables. Biotic indices determined from actual fauna are used to detect any departure from the ideal.

For the subjective approach adopted in this project, definitions of biological classes in terms of expected fauna may need to be agreed before eliciting the probability distributions that express support for each class. This definition would standardise on particular conditions, and may have to account for regional variations. The knowledge elicitation for this project has focused on the biological quality of riffles, since the fauna for these are most sensitive to pollution. Certain types of fauna may however not appear in

---

[1] A comprehensive survey of the literature on combining probability distributions has been carried out by Genest and Zidek (1986)

lowland regions that nonetheless have rivers of good biological quality. Evidence from the Northumbria and Yorkshire region suggests that biological classifications of sites by ecologists are highly correlated, implying that the assessors have a clear mental model of the classes themselves.[2]

The probabilistic knowledge elicited from the expert relates to organic pollution in riffles. There is no theoretical reason why the knowledge elicitation could not be extended to cover toxic pollution, although this complicates the idea of biological class. The probability distributions partly define the classes and also encode the degree of support for them, since this knowledge relates to organic pollution. Therefore the biological classes as defined here are essentially zones of organic enrichment, like the saprobic zones. Separate biological classes would probably be required to deal adequately with toxic pollution.

8.4 Richer knowledge structures

The structure of knowledge used for the automated classification has been essentially that of the one-layer belief network shown in Chapter 3. Here a range of observations is conditionally independent with respect to the hypotheses that cause them. The assumption of conditional independence is critical for both the Bayesian and Dempster-Shafer calculi. For the biological classification problem, the presence or absence of a particular taxon is independent of the presence or absence of another for some hypothesis H_i. Given the range of interactions that take place within a benthic community, this assumption may not always be justified.

For the Dempster-Shafer calculus the problem of independence has been considered by Lingras and Wong (1990). They introduce the idea of dependency functions in which mass assignments are adjusted if two bodies of evidence bearing on the same proposition are dependent. One side-effect of these functions is to reduce the probability mass allocated for conflicting evidence, thereby avoiding problems that arise due to normalisation in Dempster's rule. There is no clear guidance however, about how the values of dependency are determined.

Bayesian networks provide a means of encapsulating a much richer structure of knowledge and expressing dependence relationships between variables. They are now the subject of considerable research. For biological surveillance, the development of a Bayesian

[2] This evidence relates to the author's interpretation of a personal communication from Ms. Viki Hirst, of the Environment Agency, referred to in Chapter 5.

network could involve the influence of environmental variables, benthic interactions (e.g. predator/prey relationships), trophic states and so on. This would entail a major exercise in knowledge acquisition, but one that could prove fruitful.

8.5 Summary

The uncertain reasoning methods for biological classification could be incorporated as part of a decision support system for the biological surveillance of river water quality. The biological classification could be supplemented by heuristics on faunal diversity or additional benthic knowledge for the entire sample. A specification and preliminary design for such a system has been presented here, but the knowledge acquisition for and the development of such a system would entail a major project in its own right.

The chapter concluded by examining several avenues for future research into the use of uncertain reasoning methods for biological classification. Techniques for combining probability distributions from several experts were discussed. Indicator taxa would need to be identified and abundance threshold-levels agreed. The distributions themselves would be elicited with respect to agreed definitions of the biological classes.

Finally the possibility of using Bayesian networks was discussed. These offer the possibility of more complex knowledge structures involving uncertain reasoning, but require intensive knowledge acquisition.

Chapter 9

Conclusions

9.1 Introduction

In this chapter the contributions of this project as described in this thesis are reviewed, and the conclusions, drawn from the previous chapters, are presented. It concludes with a restatement of main recommendations.

9.2 Project Review

The quality of river water and the aquatic environment must be effectively monitored, reported and controlled. Biological methods have an important role to play in all three of these activities. The measurement and assessment of river water quality require an understanding of its meaning and the context in which the term is used. Throughout this thesis, the term "water quality" or "river water quality" has been understood within the context of the overall quality of the aquatic environment, with its physical, chemical and biological dimensions.

Both chemical and biological methods are required adequately to measure the parameters that define water quality. However, of the three principal media within the aquatic environment, water, particulate matter and living organisms, the last of these provide a direct indication of biological quality. Biological surveillance is therefore the natural methodology for this assessment. Of the many organisms employed for biological surveillance, the benthic macroinvertebrates associated with the river bed have several advantages that render them preeminent in this field. The techniques described in the previous chapters could however in theory be applied to other groups of biota. Despite the complexity of information inherent in biological data, the need exists to summarise and report biological quality in a form that can be readily understood by decision makers and the public. The many biotic and diversity indices that exist both in Britain and on the Continent attest to this fact.

Discrete classification schemes appear to afford a more useful means of communicating water quality, in both its chemical and biological dimensions. Accordingly, the use of a discrete biological classification system, as described in this work, is advocated for reporting biological water quality. Such systems have been used in the past by water authorities. Their use on a national level would however require considerable effort in agreeing definitions and accommodating regional variations.

The biotic indices incorporate the knowledge and experience of river biologists. For these, subjective knowledge of (for instance) the conditions preferred by particular invertebrates is directly employed in the index design. In contrast to community-structure and diversity indices, the advanced statistical techniques of RIVPACS or artificial neural-networks adopt a pattern-recognition approach for attaining objective measures of quality. Currently, statistical techniques have considerable support in that the RIVPACS methodology is being advocated for deriving an ecological quality index for national river surveys.

The impetus for this project derives from a view that the subjective and experiential knowledge of experts in biological surveillance is valuable and can be encoded for classifying river water quality. The benthic invertebrates are viewed as sensors of river water quality, whose occurrence in the range of river water qualities is characterised by uncertainty. This uncertainty can be managed by assigning numerical degrees of belief to propositions corresponding to the biological classes according to the strength of support provided by this evidence.

Techniques of uncertain reasoning provide a means by which problems characterised by uncertainty can be formulated and their information manipulated to reach rational and reproducible decisions. Of the several numerical reasoning methods, the Bayesian and Dempster-Shafer calculi provide consistent and mathematically coherent procedures for manipulating and combining evidence. This project was concerned with using and analysing these methods to both model biological data as uncertain information and to emulate the expert's ability to classify river water quality by it.

As part of the knowledge acquisition process, carried out in close cooperation with the domain expert, a group of indicator taxa deemed to be "reliable witnesses" of river water quality was selected from the large range of benthic taxa. In what is believed to be the first knowledge elicitation exercise of this kind for benthic invertebrate data, probability distributions for this group were directly elicited using several methods, culminating in the use of a graphical method developed especially for this work (Chapter 4). Abundance-levels for each indicator, including absence, were modelled as discrete sensor states, each of which had an associated probability distribution derived from the elicited distributions.

A database of benthic samples for riffle sites was constructed from extensive invertebrate distribution records in Yorkshire and classified by the expert using the proposed biological classification system (Chapter 5). Several biotic indices were calculated for each sample and compared with the expert's classifications. From the viewpoint of river water quality management, the large overlap in the ranges of the biotic indices between classes reduces their discriminatory power for detecting gradations of water quality.

Following the preliminary classification using the Bayesian calculus, an experimental programme was devised in which the Bayesian and Dempster-Shafer calculi were tested for their performance in classifying riffle sites using benthic taxa data primarily from the Yorkshire Region. To allow comparison with the expert's classification, site data was standardised as abundance levels of indicator taxa so that both the automatic classifiers and the domain expert used the same information. Decision mechanisms were designed which accounted for the ranked order of support produced by the Bayesian and Dempster-Shafer classifiers and allowed comparison with the expert's classification. By considering the expert's classification as the 'reference' class, the performance of the automatic classifiers was assessed.

The preliminary classifications also suggested that the strength of sensor evidence had implications for evidential conflict and classification performance. In emulating the expert's ability to classify using biological evidence, an uncertain reasoning classifier may have to weigh or adjust evidence appropriately before combining it, i.e. some measure of "data quality" should be applied to the raw sensor data. The effect of varying evidential strength was investigated both by adjustments to the probability distributions and by discounting part of the evidence.

The representation of sensor evidence as belief in the various propositions was investigated. Bayesian belief or so-called singleton support functions correspond to a one-to-one mapping between the probability distributions afforded by the sensor evidence and the assignment of probability mass for singleton propositions. Evidential discount can be used in the Dempster-Shafer calculus to vary the strength of sensor evidence for the singletons uniformly, to retain uncommitted belief and by that improve the effects of evidential conflict and normalisation. For Bayesian belief, evidential discount reduces conflict but distorts the sensor evidence provided by the original data.

The Dempster-Shafer calculus allows a much more flexible representation of evidence, in that belief can be assigned to the entire power set within the frame of discernment. Since this choice can be overwhelming, mass assignment is often constrained to classes of belief function. Singleton support is a variant of the Bayesian belief function in which non-zero evidential discount is applied to each focal singleton proposition. Two other classes of belief function, that of simple support and consonant belief, were chosen for investigation as the most promising alternative representations of evidence.

The performance of the Dempster-Shafer classifiers using the simple support representation gives acceptable classification performance when using present data only. Use of absent evidence, whether confirming or disconfirming resulted in poor performance. The computational behaviour of these algorithms was investigated, particularly with respect to dealing with evidential conflict.

9.3 Summary and Conclusions

The conclusions from this study are:

1. Uncertain reasoning methods afford a viable means of interpreting biological data in terms of a biological classification system. Both the Bayesian and Dempster-Shafer calculi provide mathematically coherent procedures for integrating sensor evidence. The subjectivity and uncertainty associated with biological data can be used to decide river water quality.
2. Classification of river water was carried out using evidence provided by a reference set of benthic organisms considered existing in one of four states, including absence. Methods for eliciting sensor evidence have been devised which reduce the requirement for a domain expert to specify numerical belief, which allow probability distributions to be derived for each sensor state.
3. In using sensor evidence for uncertain reasoning, attention should be paid to the weighting of the data, its representation as a belief function, the role of evidence provided by absent sensors and the behaviour of the decision algorithm with conflicting evidence.
4. The performance of Bayesian classifiers was improved by adjustments to the elicited probability distributions, since this redistribution of probability mass reduces the possibility of premature decisions before combining all available evidence.
5. The use of evidential discount provides a more controlled means of adjusting evidential strength and of managing conflict between sensor data. For the Dempster-Shafer calculus using singleton support functions, a nominal degree of evidential discount improves performance by reducing conflict and maintaining plausible belief in competing hypotheses for as long as possible.

6. The representation of belief as Bayesian or singleton support allows a performance comparison between the Bayesian and Dempster-Shafer calculi. For a nominal evidential discount, little difference was observed between the two algorithms. This results from the proximity of singleton support to Bayesian belief. In combining evidence from many sensors, singleton support functions become increasingly Bayesian. Consequently, decisions produced by Dempster-Shafer classifiers in such circumstances are often point-valued.

7. The inclusion of absent evidence for singleton support belief functions makes a significant improvement in classification performance, in spite of its lower information content. The cumulative effect of evidence induced by the absence of indicator taxa is to reinforce evidence from abundant and established taxa, while simultaneously maintaining positive support to propositions that would otherwise be vetoed during evidential combination.

8. Classification performance, when measured over the Yorkshire data set, was robust with respect to reduced decision threshold values for the Bayesian and Dempster-Shafer classifiers. If such thresholds are applied, the evidence should be ordered by sensor state such that abundant evidence is considered first, followed by established and then absent evidence. If both present and absent evidence is considered, the order of combination can be arbitrary.

9. Rare evidence was not generally assumed to affect the expert's assessment of biological quality. Results from the automatic classifiers seem to support this premise, although this may follow from the probabilistic representation of the evidence.

10. The computational behaviour of the Dempster-Shafer calculus was considered with respect to its ability to manage evidential conflict from benthic sensors. In using pair-wise applications of Dempster's rule of combination, the mass of the null-set allows the current degree of conflict between the incoming data and the combined evidence to be monitored, so that if necessary the conflicting data can be rejected. The use of empirical conflict threshold levels can however lead to poor rejection decisions. By leaving some belief uncommitted with the environment, the Dempster-Shafer calculus deals adequately with all but the most highly conflicting evidence. If conflict threshold levels are used, they should be set very high.

11. The representation of belief as simple support is more clearly non-Bayesian, and uses evidential discount more naturally than singleton support. The maximum value within the probability distributions for the established and abundant sensor states was used to

assign support for the coincident singleton proposition, and the remaining mass assigned to the environment. Barnett's computational scheme used for combining simple support evidence allows the calculation of the total weight of conflict. This parameter increases approximately linearly with the number of sensors used in the decision process.

12. Simple support functions appear to lend themselves to the use of the T-R combination rule, an alternative to Dempster's rule. However the use of this alternative in a multihypothesis space, along with the decision process used to interpret its results is not clear.

13. The use of absent evidence for simple support functions is problematic. Its representation as confirming singleton foci produces poor results, with no improvement if used to disconfirm by using evidence provided by the present sensor state. The application of variable weighting, in which the absence of higher-quality taxa was weighted less than those associated with poorer-quality waters did however reduce the serious misclassifications that occurred without weighting.

14. The classification rate using present data alone for simple support was better than when absent data was included, in contrast to singleton support and Bayesian belief. The abilities to classify on present data only, coupled with the simplicity of the simple support representation, and the availability of Barnett's scheme for rapidly combining evidence, are clearly desirable attributes.

15. Simple support functions offer the possibility of a simpler knowledge elicitation process for the indicator taxa. A domain expert could be asked to show the degree to which an indicator supports (or refutes) a single class.

16. Consonant belief functions are intuitively attractive but their use is costly from a computational viewpoint. This however may not be a problem, given the increased availability of computational power. Consonance is not preserved between evidential combination, and undesirable effects can arise from reinforcement of proposition arising from set intersections between evidence that would otherwise be conflicting. Their performance in the Dempster-Shafer classifiers is however comparable to the singleton support functions for the Yorkshire Water data set.

9.4 Closing Remarks

The principal recommendation of this project is to advocate the further use and investigation of uncertain reasoning methods for biological classification as discussed in Chapter 8.

The simplified Bayesian and Dempster-Shafer numerical reasoning schemes discussed in this project provide good decision algorithms in this context if the data is appropriately weighted to deal with evidential conflict. The Dempster-Shafer theory of evidence is a generalisation of the Bayesian method and offers more flexibility in the representation of evidence. Uncertainty in the evidence can be explicitly modelled as uncommitted belief. Thus the calculus provides an intuitively attractive framework for integrating and interpreting biological sensor data that is noisy, imprecise and often conflicting.

The acceptance of these methods for the biological classification of river water quality requires the adoption of such a classification scheme with its attendant definitions of expected fauna, an agreed reference set of indicator taxa (which may however vary from region to region), and an agreed set of sensor states. Such a selection may be guided by the frequency of occurrence of the taxa in biological surveys.

The modelling of different abundance levels, specific to each sensor, is considered an important component of biological interpretation. The role of rare evidence should be investigated further. The occurrence of taxa in very low numbers may be more significant than assumed here, arising perhaps from particular environmental stress factors with would therefore have implications for water quality. The knowledge elicitation methods devised for this project provide a model for which sensor probability distributions can be elicited or derived. Consideration should be given to combining opinions from multiple experts on both the occurrence of indicator taxa across the biological classes and their classifications of sites from benthic data.

The automatic classification software was written within the LEONARDO environment. This modular nature of this environment lends itself to expanding the biological classification software to incorporate heuristic knowledge on biological surveillance, some of which has been acquired from the domain expert while eliciting the probability distributions. This project was conceived as one part of a much larger and ambitious scheme to develop a complete decision support system for river water quality monitoring, including the location of pollution sources. The implementation of such a system could have considerable practical use as well as promoting the importance of biological methods and particularly biological classification using uncertain reasoning.

Finally, the use of Bayesian networks to model the complexity of benthic communities should be actively considered. Such a network could capture the inherent dependencies and interactions within a rich knowledge structure. As with all these methods that model uncertainty and subjectivity, the willing cooperation of an expert or group of experts is required.

Appendix A1

National Water Council Classification Scheme for Water Quality in Rivers and Canals

Description	Class	Current Potential Use
Good Quality	1a	Water of high quality suitable for potable supply abstractions; game or other high class fisheries; high amenity value.
	1b	Water of less high quality than Class 1a but usable for substantially the same purpose.
Fair Quality	2	Water suitable for potable supply after advanced treatment; supporting reasonably good coarse fisheries; moderate ammenity value.
Poor Quality	3	Waters which are polluted to an extent that fish are absent or only sporadically present; may be used as a low grade industrial abstraction; considerable potential use for further use if cleaned up.
Bad Quality	4	Waters which are grossly polluted and are likely to cause nuisance.

Appendix A2
Elicited Probability Distributions

Table A2.1 Heights of histograms for each indicator taxon in two states: Present and Abundant

Group	Code No.	Taxon	B1a	B1b	B2	B3	B4	P(e|pref)
C	3120201	Polycelis nigra	2	11	15	2	0	0.100
			0	11	15	1	0	0.075
C	3130101	Dendrocoelum lacteum	0	3	14	2	0	0.100
			0	4	17	1	0	0.010
C	13040301	Potamopyrgus jenkinsi	14	14	10	5	0	0.200
			10	19	5	0	0	0.150
C	13040501	Bithynia tentaculata	0	9	20	2	0	0.100
			0	7	21	0	0	0.020
A	13070201	Lymnaea peregra	5	25	35	30	5	0.500
			0	20	50	28	0	0.125
C	13090300	Planorbis spp.	0	9	17	5	0	0.100
			0	5	17	11	0	0.020
C	13100201	Ancylus fluviatilis	11	11	11	5	0	0.200
			5	11	18	3	0	0.150
C	14030100	Sphaerium spp.	0	8	21	17	0	0.100
			0	4	28	23	0	0.075
C	14030200	Pisidium spp.	6	19	9	4	0	0.100
			0	26	6	0	0	0.075
A[1]	16030000	TUBIFICIDAE	0	0	4	19	32	0.900
			0	0	0	6	39	0.700
B	16060000	LUMBRICULIDAE	0	20	20	20	0	0.700
			0	5	27	9	0	0.200
B	17020300	Glossiphonia spp.	0	0	40	16	0	0.400
			0	0	44	10	0	0.130
B	17020501	Helobdella stagnalis	0	0	17	46	0	0.400
			0	0	10	50	0	0.040
A	17040102	Erpobdella octoculata	0	0	17	26	6	0.700
			0	0	10	34	2	0.070
C	19000000	HYDRACARINA	9	9	9	0	0	0.200
			10	22	10	0	0	0.150
A	28030101	Asellus aquaticus	0	2	20	14	2	0.900
			0	0	28	12	0	0.630
A	28070305	Gammarus pulex	11	16	10	0	0	0.950
			5	25	7	0	0	0.700
B	30020105	Baetis rhodani	22	22	22	16	0	0.750
			0	31	25	3	0	0.380
B	30030100	Rhithrogena spp.	25	17	0	0	0	0.500
			28	10	0	0	0	0.130

Table A2.1 Heights of histograms for each indicator taxon in two states: Present and Abundant

Group	Code No.	Taxon	B1a	B1b	B2	B3	B4	P(e|pref)
C	30030200	Heptagenia spp.	26	8	0	0	0	0.200
			27	0	0	0	0	0.100
B	30030400	Ecdyonurus spp.	20	36	0	0	0	0.500
			12	38	0	0	0	0.130
C	30050101	Ephemerella ignita	3	16	16	4	0	0.500
			0	21	14	0	0	0.375
B	30080200	Caenis spp.	13	31	0	0	0	0.500
			16	43	0	0	0	0.200
C	31020202	Amphinemura sulcicollis	15	9	4	0	0	0.200
			22	3	0	0	0	0.100
A[2]	31030100	Leuctra spp.	20	14	4	0	0	0.500
			16	16	4	0	0	0.100
C	31050401	Isoperla grammatica	25	15	2	0	0	0.200
			25	4	0	0	0	0.100
B	35010000	HALIPLIDAE	0	20	20	0	0	0.100
			0	20	20	0	0	0.010
C	35030000	DYTISCIDAE	0	21	21	4	0	0.200
			0	23	23	0	0	0.040
C	35110000	ELMINTHIDAE	16	16	5	0	0	0.200
			12	23	7	0	0	0.100
B	36010101	Sialis lutaria	0	0	18	26	0	0.150
			0	0	10	40	0	0.030
A[3]	38010100	Rhyacophila spp.	25	30	8	0	0	0.150
			17	34	12	0	0	0.015
C	38010200	Glossosoma spp.	23	14	0	0	0	0.200
			24	5	0	0	0	0.150
C	38010300	Agapetus spp.	28	17	0	0	0	0.200
			28	6	0	0	0	0.150
C	38030000	POLYCENTROPODIDAE	14	14	3	0	0	0.200
			11	19	1	0	0	0.100
A	38050111	Hydropsyche angustipennis	0	9	29	14	0	0.250
			0	3	38	10	0	0.125
B	38050112	Other HYDROPSYCHIDAE	14	32	0	0	0	0.250
			8	50	0	0	0	0.050
C	38060000	HYDROPTILIDAE	11	19	10	0	0	0.200
			9	21	17	0	0	0.100
C	38080000	LIMNEPHILIDAE	19	19	10	0	0	0.200
			18	27	4	0	0	0.150
C	39111111	CERATOPOGONIDAE	18	18	14	3	0	0.200
			20	20	8	1	0	0.100
A	40140215	Chironomus riparius	0	0	2	23	23	0.900
			0	0	0	21	27	0.630

Table A2.1 Heights of histograms for each indicator taxon in two states: Present and Abundant

Group	Code No.	Taxon	B1a	B1b	B2	B3	B4	P(e|pref)
A	40150510	Simulium ornatum	0	0	26	24	0	0.100
			0	0	26	24	0	0.050

Notes:
P (e|pref) refers to probability of finding taxon in sate for its preferred class.
First row for each taxon refers to 'Present' state, second to 'Abundant'.
See Table 4.4 for notes on 'Group'.
Code No. refers to codes mostly taken from Maitland (1977).

Appendix A3
Derived Probability Data

Table A3.1 Summary Data Derived from Heights of Elicited Histograms

Indicator Taxon	Class	Established			Abundant			Absent		
		P(H|e)	P(e|H)	Ls	P(H|e)	P(e|H)	Ls	*P(H|e)*	*P(e|H)*	*Ls*
Polycelis nigra	B1a	0.21	0.01	1.03	0.00	0.00	0.00	0.21	0.99	1.03
	B1b	0.28	0.02	1.57	0.41	0.06	2.75	0.19	0.93	0.96
	B2	0.38	0.03	2.50	0.56	0.08	5.00	0.19	0.90	0.92
	B3	0.13	0.01	0.15	0.04	0.01	0.15	0.21	0.99	1.03
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.05
Dendrocoelum lacteum	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.04
	B1b	0.16	0.02	0.74	0.18	0.00	0.89	0.20	0.98	1.01
	B2	0.73	0.09	10.98	0.77	0.01	13.60	0.19	0.90	0.91
	B3	0.11	0.01	0.13	0.05	0.00	0.19	0.20	0.99	1.02
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.04
Potamopyrgus jenkinsi	B1a	0.35	0.12	2.15	0.29	0.08	1.67	0.18	0.80	0.89
	B1b	0.14	0.05	0.68	0.56	0.15	5.07	0.18	0.80	0.89
	B2	0.30	0.10	1.71	0.15	0.04	0.69	0.20	0.86	0.97
	B3	0.21	0.07	0.26	0.00	0.00	0.00	0.21	0.93	1.07
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.23	1.00	1.18
Bithynia tentaculata	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.04
	B1b	0.30	0.04	1.70	0.25	0.01	1.33	0.20	0.96	0.98
	B2	0.62	0.08	6.62	0.75	0.02	12.00	0.19	0.90	0.91
	B3	0.08	0.01	0.08	0.00	0.00	0.00	0.20	0.99	1.03
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.04
Lymnaea peregra	B1a	0.06	0.07	0.26	0.00	0.00	0.00	0.26	0.93	1.41
	B1b	0.26	0.31	1.40	0.20	0.05	1.03	0.18	0.64	0.88
	B2	0.32	0.38	1.86	0.51	0.13	4.17	0.14	0.50	0.65
	B3	0.30	0.36	0.43	0.29	0.07	1.60	0.16	0.57	0.76
	B4	0.06	0.07	0.06	0.00	0.00	0.00	0.26	0.93	1.41
Planorbis spp.	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.05
	B1b	0.33	0.05	1.95	0.15	0.01	0.71	0.20	0.95	0.98
	B2	0.56	0.08	5.04	0.52	0.02	4.25	0.19	0.90	0.92
	B3	0.11	0.02	0.13	0.33	0.01	2.00	0.20	0.97	1.01
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.05
Ancylus fluviatilis	B1a	0.41	0.16	2.82	0.14	0.04	0.63	0.19	0.80	0.91
	B1b	0.28	0.11	1.58	0.30	0.09	1.69	0.19	0.80	0.91
	B2	0.13	0.05	0.60	0.49	0.15	3.79	0.19	0.80	0.91
	B3	0.17	0.07	0.21	0.08	0.03	0.35	0.21	0.91	1.07
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.23	1.00	1.21
Sphaerium spp.	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.06
	B1b	0.38	0.03	2.47	0.07	0.01	0.31	0.20	0.96	1.01
	B2	0.35	0.03	2.14	0.51	0.08	4.15	0.19	0.90	0.93
	B3	0.27	0.02	0.37	0.42	0.06	2.88	0.19	0.92	0.95
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.06

Table A3.1 Summary Data Derived from Heights of Elicited Histograms

Indicator Taxon	Class	Established			Abundant			Absent		
		P(H|e)	P(e|H)	Ls	P(H|e)	P(e|H)	Ls	*P(H|e)*	*P(e|H)*	*Ls*
Pisidium spp.	B1a	0.29	0.03	1.66	0.00	0.00	0.00	0.20	0.97	1.01
	B1b	0.23	0.03	1.21	0.81	0.08	17.33	0.19	0.90	0.92
	B2	0.28	0.03	1.55	0.19	0.02	0.92	0.20	0.95	0.99
	B3	0.20	0.02	0.24	0.00	0.00	0.00	0.20	0.98	1.02
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.05
TUBIFICIDAE	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.29	1.00	1.63
	B1b	0.00	0.00	0.00	0.00	0.00	0.00	0.29	1.00	1.63
	B2	0.15	0.11	0.72	0.00	0.00	0.00	0.26	0.89	1.38
	B3	0.58	0.43	1.37	0.13	0.11	0.62	0.13	0.47	0.62
	B4	0.27	0.20	0.37	0.87	0.70	26.00	0.03	0.10	0.12
LUMBRICULIDAE	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.34	1.00	2.11
	B1b	0.37	0.66	2.34	0.12	0.04	0.56	0.10	0.30	0.46
	B2	0.28	0.50	1.54	0.66	0.20	7.71	0.10	0.30	0.46
	B3	0.35	0.63	0.54	0.22	0.07	1.13	0.10	0.30	0.46
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.34	1.00	2.11
Glossiphonia spp.	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.23	1.00	1.16
	B1b	0.00	0.00	0.00	0.00	0.00	0.00	0.23	1.00	1.16
	B2	0.67	0.27	8.28	0.81	0.13	17.60	0.14	0.60	0.63
	B3	0.33	0.13	0.48	0.19	0.03	0.91	0.19	0.84	0.93
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.23	1.00	1.16
Helobdella stagnalis	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.16
	B1b	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.16
	B2	0.28	0.14	1.55	0.17	0.01	0.80	0.19	0.85	0.95
	B3	0.72	0.36	2.57	0.83	0.04	20.00	0.13	0.60	0.62
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.16
Erpobdella octoculata	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.27	1.00	1.49
	B1b	0.00	0.00	0.00	0.00	0.00	0.00	0.27	1.00	1.49
	B2	0.36	0.44	2.22	0.22	0.02	1.11	0.15	0.54	0.69
	B3	0.51	0.63	1.06	0.74	0.07	11.33	0.08	0.30	0.35
	B4	0.13	0.16	0.15	0.04	0.00	0.18	0.23	0.84	1.18
HYDRACARINA	B1a	0.42	0.13	2.90	0.24	0.07	1.25	0.18	0.80	0.89
	B1b	0.16	0.05	0.76	0.52	0.15	4.40	0.18	0.80	0.89
	B2	0.42	0.13	2.90	0.24	0.07	1.25	0.18	0.80	0.89
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.23	1.00	1.18
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.23	1.00	1.18
Asellus aquaticus	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.30	1.00	1.75
	B1b	0.11	0.09	0.50	0.00	0.00	0.00	0.28	0.91	1.53
	B2	0.33	0.27	2.00	0.70	0.63	9.33	0.03	0.10	0.13
	B3	0.44	0.36	0.80	0.30	0.27	1.71	0.11	0.37	0.51
	B4	0.11	0.09	0.13	0.00	0.00	0.00	0.28	0.91	1.53
Gammarus pulex	B1a	0.44	0.51	3.17	0.14	0.14	0.63	0.12	0.35	0.56
	B1b	0.22	0.25	1.10	0.68	0.70	8.33	0.02	0.05	0.07
	B2	0.34	0.40	2.08	0.19	0.20	0.93	0.14	0.41	0.68
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.36	1.00	2.22
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.36	1.00	2.22

Table A3.1 Summary Data Derived from Heights of Elicited Histograms

Indicator Taxon	Class	Established			Abundant			Absent		
		P(H|e)	P(e|H)	Ls	P(H|e)	P(e|H)	Ls	*P(H|e)*	*P(e|H)*	*Ls*
Baetis rhodani	B1a	0.36	0.75	2.27	0.00	0.00	0.00	0.11	0.25	0.51
	B1b	0.18	0.37	0.87	0.53	0.38	4.43	0.11	0.25	0.51
	B2	0.21	0.44	1.09	0.42	0.31	2.94	0.11	0.25	0.51
	B3	0.25	0.51	0.33	0.05	0.04	0.21	0.21	0.45	1.04
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.45	1.00	3.32
Rhithrogena spp.	B1a	0.56	0.37	5.04	0.74	0.13	11.20	0.12	0.50	0.55
	B1b	0.44	0.29	3.17	0.26	0.05	1.43	0.16	0.66	0.75
	B2	0.00	0.00	0.00	0.00	0.00	0.00	0.24	1.00	1.27
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.24	1.00	1.27
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.24	1.00	1.27
Heptagenia spp.	B1a	0.62	0.10	6.50	1.00	0.10	0.00	0.17	0.80	0.81
	B1b	0.38	0.06	2.46	0.00	0.00	0.00	0.20	0.94	0.99
	B2	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.07
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.07
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.07
Ecdyonurus spp.	B1a	0.39	0.24	2.56	0.24	0.04	1.26	0.17	0.72	0.83
	B1b	0.61	0.37	6.25	0.76	0.13	12.67	0.12	0.50	0.54
	B2	0.00	0.00	0.00	0.00	0.00	0.00	0.24	1.00	1.24
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.24	1.00	1.24
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.24	1.00	1.24
Ephemerella ignita	B1a	0.16	0.09	0.75	0.00	0.00	0.00	0.24	0.91	1.26
	B1b	0.21	0.13	1.07	0.60	0.38	6.00	0.13	0.50	0.61
	B2	0.42	0.25	2.91	0.40	0.25	2.67	0.13	0.50	0.61
	B3	0.21	0.13	0.27	0.00	0.00	0.00	0.23	0.88	1.20
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.26	1.00	1.44
Caenis spp.	B1a	0.31	0.14	1.80	0.27	0.07	1.49	0.18	0.79	0.90
	B1b	0.69	0.30	8.87	0.73	0.20	10.75	0.12	0.50	0.53
	B2	0.00	0.00	0.00	0.00	0.00	0.00	0.23	1.00	1.22
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.23	1.00	1.22
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.23	1.00	1.22
Amphinemura sulcicollis	B1a	0.39	0.10	2.50	0.88	0.10	29.33	0.17	0.80	0.84
	B1b	0.41	0.11	2.77	0.12	0.01	0.55	0.19	0.88	0.94
	B2	0.21	0.05	1.03	0.00	0.00	0.00	0.20	0.95	1.03
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.10
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.10
Leuctra spp.	B1a	0.55	0.40	4.92	0.44	0.10	3.20	0.12	0.50	0.56
	B1b	0.34	0.25	2.11	0.44	0.10	3.20	0.16	0.65	0.76
	B2	0.10	0.08	0.46	0.11	0.03	0.50	0.22	0.90	1.14
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.25	1.00	1.31
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.25	1.00	1.31
Isoperla grammatica	B1a	0.45	0.10	3.33	0.86	0.10	25.00	0.17	0.80	0.83
	B1b	0.47	0.10	3.59	0.14	0.02	0.64	0.19	0.88	0.93
	B2	0.07	0.02	0.31	0.00	0.00	0.00	0.21	0.98	1.07
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.09
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.09

Table A3.1 Summary Data Derived from Heights of Elicited Histograms

Indicator Taxon	Class	Established			Abundant			Absent		
		P(H|e)	P(e|H)	Ls	P(H|e)	P(e|H)	Ls	*P(H|e)*	*P(e|H)*	*Ls*
HALIPLIDAE	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.05
	B1b	0.50	0.09	4.00	0.50	0.01	4.00	0.19	0.90	0.92
	B2	0.50	0.09	4.00	0.50	0.01	4.00	0.19	0.90	0.92
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.05
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.05
DYTISCIDAE	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.12
	B1b	0.45	0.16	3.23	0.50	0.04	4.00	0.18	0.80	0.85
	B2	0.45	0.16	3.23	0.50	0.04	4.00	0.18	0.80	0.85
	B3	0.11	0.04	0.12	0.00	0.00	0.00	0.21	0.96	1.07
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.12
ELMINTHIDAE	B1a	0.53	0.15	4.48	0.29	0.05	1.60	0.18	0.80	0.86
	B1b	0.36	0.10	2.22	0.55	0.10	4.84	0.18	0.80	0.86
	B2	0.11	0.03	0.52	0.17	0.03	0.80	0.21	0.94	1.04
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.13
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.13
Sialis lutaria	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.07
	B1b	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.07
	B2	0.45	0.10	3.21	0.20	0.01	1.00	0.19	0.90	0.93
	B3	0.55	0.12	1.25	0.80	0.03	16.00	0.18	0.85	0.87
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.07
Rhyacophila spp.	B1a	0.41	0.12	2.77	0.27	0.01	1.48	0.19	0.88	0.92
	B1b	0.47	0.14	3.55	0.54	0.02	4.69	0.18	0.85	0.89
	B2	0.12	0.03	0.55	0.19	0.01	0.94	0.20	0.96	1.03
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.09
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.09
Glossosoma spp.	B1a	0.36	0.05	2.21	0.83	0.15	19.20	0.17	0.80	0.83
	B1b	0.64	0.09	7.24	0.17	0.03	0.83	0.19	0.88	0.92
	B2	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.09
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.09
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.09
Agapetus spp.	B1a	0.36	0.05	2.24	0.82	0.15	18.67	0.17	0.80	0.83
	B1b	0.64	0.09	7.14	0.18	0.03	0.86	0.19	0.88	0.92
	B2	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.09
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.09
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.09
POLYCENTROPODIDAE	B1a	0.51	0.14	4.13	0.35	0.06	2.20	0.18	0.80	0.85
	B1b	0.36	0.10	2.23	0.61	0.10	6.33	0.18	0.80	0.85
	B2	0.13	0.04	0.62	0.03	0.01	0.13	0.21	0.96	1.06
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.12
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.12
Hydropsyche angustipennis	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.13
	B1b	0.24	0.07	1.27	0.06	0.01	0.25	0.20	0.92	1.02
	B2	0.45	0.13	3.22	0.75	0.13	11.69	0.16	0.75	0.79
	B3	0.31	0.09	0.46	0.20	0.03	0.98	0.19	0.88	0.96
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.13

Table A3.1 Summary Data Derived from Heights of Elicited Histograms

Indicator Taxon	Class	Established P(H|e)	Established P(e|H)	Established Ls	Abundant P(H|e)	Abundant P(e|H)	Abundant Ls	Absent P(H|e)	Absent P(e|H)	Absent Ls
Other HYDROPSYCHIDAE	B1a	0.34	0.10	2.03	0.14	0.01	0.64	0.19	0.89	0.95
	B1b	0.66	0.20	7.89	0.86	0.05	25.00	0.16	0.75	0.77
	B2	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.10
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.10
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.10
HYDROPTILIDAE	B1a	0.37	0.07	2.35	0.19	0.04	0.95	0.19	0.88	0.96
	B1b	0.51	0.10	4.11	0.45	0.10	3.23	0.17	0.80	0.85
	B2	0.12	0.02	0.56	0.36	0.08	2.27	0.20	0.89	0.97
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.12
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.12
LIMNEPHILIDAE	B1a	0.43	0.10	3.01	0.37	0.10	2.32	0.18	0.80	0.87
	B1b	0.21	0.05	1.09	0.55	0.15	4.91	0.18	0.80	0.87
	B2	0.36	0.08	2.21	0.08	0.02	0.36	0.20	0.89	0.99
	B3	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.14
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.00	1.14
CERATOPOGONIDAE	B1a	0.29	0.10	1.64	0.41	0.10	2.76	0.18	0.80	0.89
	B1b	0.29	0.10	1.64	0.41	0.10	2.76	0.18	0.80	0.89
	B2	0.34	0.12	2.02	0.16	0.04	0.78	0.19	0.84	0.95
	B3	0.08	0.03	0.09	0.02	0.01	0.08	0.22	0.97	1.12
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.23	1.00	1.17
Chironomus riparius	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.32	1.00	1.89
	B1b	0.00	0.00	0.00	0.00	0.00	0.00	0.32	1.00	1.89
	B2	0.10	0.08	0.46	0.00	0.00	0.00	0.30	0.92	1.68
	B3	0.54	0.41	1.18	0.44	0.49	3.11	0.03	0.10	0.13
	B4	0.36	0.27	0.55	0.56	0.63	5.14	0.03	0.10	0.13
Simulium ornatum	B1a	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.05
	B1b	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.05
	B2	0.52	0.05	4.33	0.52	0.05	4.33	0.19	0.90	0.92
	B3	0.48	0.05	0.92	0.48	0.05	3.69	0.19	0.91	0.93
	B4	0.00	0.00	0.00	0.00	0.00	0.00	0.21	1.00	1.05

Appendix A4

Barnett's Scheme for Combining Simple Support Functions

This scheme can be used in Dempster-Shafer reasoning if the evidence can be represented as support either for a singleton hypothesis or its complement. According to Barnett (1981) direct implementation of Dempster's rule for arbitrary belief functions is not feasible since the algorithmic complexity is of exponential order. Barnett's scheme can be carried out in linear time.

If s_{ij} is degree of support for the singleton hypothesis H_i from a sensor j, then the total support from $j = 1$ to k sensors for that proposition is:

$$f_i = 1 - \prod_{j=1}^{k} (1 - s_{ij}) \tag{A4.1}$$

This follows from the application of Dempster's rule. If s_1 is the basic probability number assigned to a singleton hypothesis, then $1 - s_1$ is assigned to the environment Θ. If s_2 is the degree of support from a second *bpa* for the same hypothesis, application of Dempster's rule results in the following:

	$i\ (s_2)$	$\Theta\ (1 - s_2)$
$i\ (s_1)$	$i\ (s_1 s_2)$	$i\ (s_1 - s_1 s_2)$
$\Theta\ (1 - s_1)$	$i\ (s_2 - s_1 s_2)$	$\Theta\ (1 - s_1 - s_2 - s_1 s_2)$

so that

$$\begin{aligned} m_1 \oplus m_2(H_i) &= s_1 s_2 + s_1 - s_1 s_2 + s_2 - s_1 s_2 \\ &= 1 - (1 - s_1)(1 - s_2) \end{aligned} \tag{A4.2}$$

Note that support for $\Theta = 1 - f_i$. If sensors $j = k+1$ to $k+l$ provide evidence against hypothesis H_i (or support for $\neg H_i$) then:

$$a_i = 1 - \prod_{j=k+1}^{k+l} (1 - s_{ij}) \tag{A4.3}$$

represents the total weight of evidence against the singleton. The support for Θ from these sensors is $1 - a_i$. The simple evidence function e_i for hypothesis H_i is then evaluated from the orthogonal sum of f_i and a_i, according to equation (7.3). Thus there are $n = 5$ simple evidence functions for the frame of biological classes.

The basic probability numbers arising from this combination are $p_i = e_i(\{H_i\}) = K_i f_i (1 - a_i)$, $c_i = e_i(\{\neg H_i\}) = K_i a_i(1 - f_i)$ and $r_i = e_i(\Theta) = K_i(1 - a_i)(1 - f_i)$, where $K_i = (1 - a_i f_i)^{-1}$ represents the degree of conflict between evidence for and against H_i. To see this, consider that $f_i = s_1$ is the total support for H_i and $a_i = s_2$ is the total support for $\neg H_i$. Combination of these via Dempster's rule then results in:

	$\neg i\ (s_2)$	$\Theta\ (1 - s_2)$
$i\ (s_1)$	$\varnothing\ (s_1 s_2)$	$i\ (s_1 (1 - s_2))$
$\Theta\ (1 - s_1)$	$\neg i\ (s_2 (1 - s_1))$	$\Theta\ ((1 - s_1)(1 - s_2))$

Because of mass in the null set, each probability number is divided by the normalisation factor $K_i = 1 - s_1 s_2 = 1 - f_i a_i$ so that for instance the basic probability number for H_i is $s_1(1 - s_2)/(1 - s_1 s_2)$, i.e. $p_i = K_i f_i\,(1 - a_i)$. The results for c_i and r_i follow in the same way.

The result of this combination is n basic probability assignments, one for each singleton. These must now be combined using Dempster's rule to form an overall belief function for each singleton. Dempster's rule is given by equations **(3.22)** and **(3.23)**, in which the normalisation factor $K^{-1} = 1 - \kappa$, where κ represents the degree of conflict between two basic probability assignments. Barnett derives a general expression for K^{-1} in terms of the basic probability numbers p_i, c_i, r_i and the overall belief in each singleton:

$$Bel(\{H_i\}) = K\left[p_i \prod_{j \neq i} d_j + r_i \prod_{j \neq i} c_j \right] \tag{A4.4}$$

where $d_i = c_i + r_i$ and

$$K^{-1} = \left[\prod_{all\ i} d_i \right]\left[1 + \sum_{all\ i} p_i / d_i \right] - \prod_{all\ i} c_i \tag{A4.5}$$

For a subset A of Θ for which $|A| > 1$ (e.g. $\neg\{Bla\}$) then

$$Bel(A) = K\left(\left[\prod_{all\ i} d_i \right]\left[\sum_{i \in A} p_i / d_i \right] + \left[\prod_{i \notin A} c_i \right]\left[\prod_{i \in A} d_i \right] - \prod_{all\ i} c_i \right) \tag{A4.6}$$

Barnett provides a complete derivation. To see how **(A4.4)** and **(A4.5)** arise consider the combination of the simple evidence functions e_i where $i = 1,2$:

	p_2	c_2	r_2
p_1	$\varnothing\ (p_1 p_2)$	$1\ (p_1 c_2)$	$1\ (p_1 r_2)$
c_1	$2\ (p_2\, c_1)$	$\varnothing\ (c_1 c_2)$	$\neg 1\ (c_1 r_2)$
r_1	$2\ (p_2\, r_1)$	$\neg 2\ (c_2 r_1)$	$\Theta\ (r_1 r_2)$

Since K^{-1} is the sum of masses whose intersection is <u>not</u> the null set, it follows from the above table that:

$$K^{-1} = p_1(c_2 + r_2) + p_2(c_1 + r_1) + c_1r_2 + c_2r_1 + r_1r_2$$
$$= p_1d_2 + p_2d_1 + c_1r_2 + c_2r_1 + r_1r_2$$
$$= d_1d_2 \left(1 + \frac{p_1}{d_1} + \frac{p_2}{d_2} \right) - c_1c_2 \qquad \textbf{(A4.7)}$$

for $d_i = c_i + r_i$. This has the form of **(A4.5)**. From the table the evidence for $\{1\}$ after combination is $K\,[p_1c_2 + p_1r_2 + r_1c_2] = K\,[p_1d_2 + r_1c_2]$ after normalisation. It should be possible to see how pair-wise combinations of the e_i would lead to **(A4.4)**.

If the simple support functions all confirm the hypotheses H_i, then the c_i are equal to zero and K^{-1} reduces to:

$$K^{-1} = \left(\prod_i r_i \right)\left(1 + \sum_i p_i/r_i \right) \qquad \textbf{(A4.8)}$$

and

$$Bel(\{H_i\}) = K\,p_i \prod_{j \neq i} r_j \qquad \textbf{(A4.9)}$$

Writing

$$\sigma = 1 + \sum_{all\ i} p_i/r_i \qquad \textbf{(A4.10)}$$

then

$$Bel(\{H_i\}) = \frac{p_i}{r_i\sigma} \qquad \textbf{(A4.11)}$$

Similarly **(A4.6)** reduces to

$$Bel(\neg\{H_i\}) = \frac{\sigma - 1}{\sigma} - \frac{p_i}{r_i\sigma} \qquad \textbf{(A4.12)}$$

The width of the evidential interval $[Bel, Pls]$ for a given hypothesis H_i is then $Pls(\{H_i\}) - Bel(\{H_i\}) = 1 - Bel(\neg\{H_i\}) - Bel(\{H_i\}) = 1/\sigma$, which holds for all singletons. Hence the evidential width is identical for all singletons if the simple support evidence is all confirming. This is not true if some of the evidence is disconfirming.

Appendix A5

T-R Combination Rule for Interval-valued Evidence

This rule has been devised by Chang and Kashyap (1990) for combining evidence which can be expressed as an interval. They claim it has properties which make it particularly suitable for conflicting evidence. The authors test the rule by combining both conflicting and non-conflicting sources, and compare its performance with Dempster's rule and the "interval Bayes' rule". The combination rule was used for the simple support representation of the benthic sensor evidence.

For an evidential interval $[a,b]$ $(0 \leq a \leq b \leq 1)$ regarding an hypothesis H, a discrimination measure Δ is defined such that if the measure is positive, H is regarded as true if a decision must be made while $\neg H$ is regarded as true if the measure is negative. Support for H varies from a to b while support for $\neg H$ varies from (1-b) to (1-a). Defining Δ to be (lower limit of support for H) - (lower limit of support for $\neg H$, we have:

$$\Delta = a - (1 - b) = a + b + 1 \tag{A5.1}$$

Two pieces of evidence $[a,b]$ and $[c,d]$ are conflicting if their discrimination measures are of the opposite signs.

As described in Chapter 8 the interval $[a,b]$ can be represented as a vector in a two-dimensional co-ordinate system, bounded by a triangle of co-ordinates (0,0), (0,1),(1,1). The vector (a,b) in the triangular region is mapped by a geometric transform to a vector (u,v) in a rectangular region, shown in Figure A5.1, such that u is equivalent to the discrimination measure. Intervals in the rectangular region are combined using functions which preserve associativity and commutativity, and the mapped back to a point (e,f) in the original triangular space. Then $[e,f]$ is the resulting interval. The transforms and combination functions are stated here without proof. See Chang and Kashyap for a complete derivation.

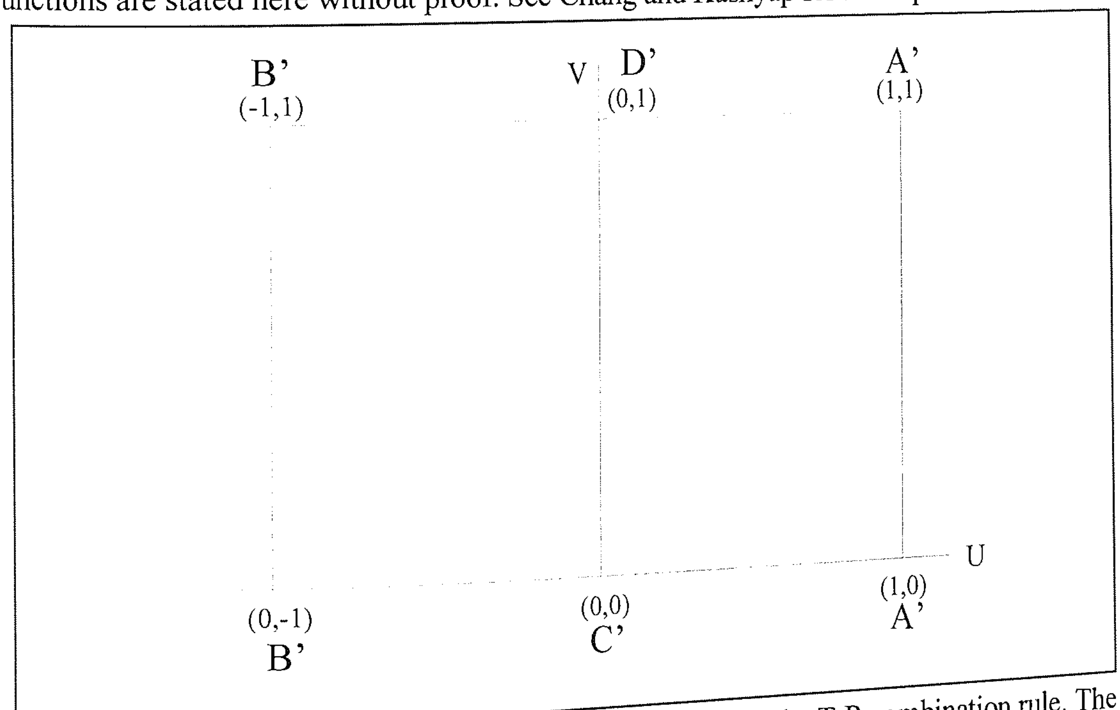

Figure A5.1 The rectangular region used for formulation of the T-R combination rule. The labels A',B',C' and D' refer to mappings from the triangular region (A,B,C,D) to their equivalent in the rectangular region.

Consider two intervals $[a_1,b_1]$, $[a_2,b_2]$. $[a_1,b_1]$ transforms to $[u_1,v_1]$ as follows:-

$$u_1 = (a_1 + b_1 - 1)^{\frac{2 - a_1 - b_1}{2 - 2a_1}}, \qquad \text{if } a_1 \geq 1 - b_1$$
$$= -(1 - a_1 - b_1)^{\frac{a_1 + b_1}{2b_1}}, \qquad \text{if } a_1 < 1 - b_1 \qquad \textbf{(A5.2)}$$

and

$$v_1 = \frac{2-2b_1}{2-a_1-b_1}, \qquad \text{if } a_1 \geq 1 - b_1$$
$$= \frac{2a_1}{a_1+b_1}, \qquad \text{if } a_1 < 1 - b_1 \qquad \textbf{(A5.3)}$$

Similarly $[a_2,b_2]$ is mapped to $[u_2,v_2]$. The components are then combined:

$$u = \frac{u_1 + u_2}{1 + u_1 u_2}$$
$$v = \frac{1}{1 + \frac{(1 - v_1)(1 - v_2)}{[(v_1 - v_1 v_2)^4 + (v_2 - v_1 v_2)^4]^{1/4}}} \qquad \textbf{(A5.4)}$$

The (u,v) are then mapped back into the original (a,b):

$$[\, e, f \,]$$
$$= [\frac{u^{(2-v)} + 1 - (1-v)(1-u^{(2-v)})}{2}, \frac{u^{(2-v)} + 1 + (1-v)(1-u^{(2-v)})}{2}], \qquad \text{if } u \geq 0$$
$$= [\frac{-|u|^{(2-v)} + 1 - (1-v)(1-|u|^{(2-v)})}{2}, \frac{-|u|^{(2-v)} + 1 + (1-v)(1-|u|^{(2-v)})}{2}], \qquad \text{if } u < 0 \qquad \textbf{(A5.5)}$$

so that the width of the interval is

$$|f - e| = (1 - v)(1 - |u|^{(2-v)}) \qquad \textbf{(A5.6)}$$

If the evidence is conflicting, the uncertainty (i.e. the evidential width) of the final result is greater than those of the original components. For non-conflicting evidence, the width reduces.

Appendix A6

Index of Experiments

Classification experiments referred to in the thesis are listed here, along with the Table or Figure in which their results are presented.

Test Description	**TestID**	**Thesis Reference**	**Decision Algorithm/ Belief Representation**	**Conflict Threshold/ Ordering/ Decision Threshold**	**Abundant/ Established/ Rare/ Absent Data Quality**
Bayesian belief- Original distributions	T200_R2	Tables 6.5, 6.8	Bayes Bayesian	Infinite Sensor-state Definite	Certain Certain Ignore Certain
Bayesian belief-Area-adjusted distributions	T201_R2	Tables 6.6, 6.8	Bayes Bayesian	Infinite Sensor-state Definite	Certain Certain Ignore Certain
Variations in evidential strength	T221	Table 6.10	Bayes Singleton	Infinite Sensor-state Definite	Certain Certain Ignore Certain
Variations in evidential strength	T222	Table 6.10	Dempster-Shafer Singleton	Infinite Sensor-state	Certain Certain Ignore Certain
Variations in evidential strength	T223	Tables 6.10, 6.13	Bayes Singleton	Infinite Sensor-state Definite	High High Ignore Good
Variations in evidential strength	T224	Tables 6.10, 6.13, 6.17	Dempster-Shafer Singleton	Infinite Sensor-state Definite	High High Ignore Good
Variations in evidential strength	T225	Tables 6.10, 6.13,6.14	Bayes Singleton	Infinite Sensor-state Definite	High High Ignore Fair
Variations in evidential strength	T226	Tables 6.10, 6.13,6.14	Dempster-Shafer Singleton	Infinite Sensor-state Definite	High High Ignore Fair
Variations in evidential strength	T227	Tables 6.10, 6.13	Bayes Singleton	Infinite Sensor-state Definite	High High Ignore Poor
Variations in evidential strength	T228	Tables 6.10, 6.13	Dempster-Shafer Singleton	Infinite Sensor-state Definite	High High Ignore Poor

Test Description	TestID	Thesis Reference	Decision Algorithm/ Belief Representation	Conflict Threshold/ Ordering/ Decision Threshold	Abundant/ Established/ Rare/ Absent Data Quality
Variations in evidential strength	T229	Table 6.10	Bayes Singleton	Infinite Sensor-state Definite	High High Ignore Ignore
Variations in evidential strength	T230	Tables 6.10,6.11, 6.12, 6.13, 6.14	Dempster-Shafer Singleton	Infinite Sensor-state Definite	High High Ignore Ignore
Variations in evidential strength	T234	Table 6.15	Dempster-Shafer Singleton	Infinite Sensor-state Definite	High Good Ignore Ignore
Variations in evidential strength	T236	Table 6.11	Dempster-Shafer Singleton	Infinite Sensor-state Definite	High Fair Ignore Ignore
Variations in evidential strength	T2361	Table 6.11	Dempster-Shafer Singleton	Infinite Sensor-state Definite	High Poor Ignore Ignore
Variations in evidential strength	T2381	Table 6.16	Dempster-Shafer Singleton	Infinite Sensor-state Definite	High High Fair Ignore
Variations in evidential strength	T238	Table 6.16	Dempster-Shafer Singleton	Infinite Sensor-state Definite	High High Poor Ignore
Decision threshold	T241TDSC	Figure 6.1, Table 6.21	Dempster-Shafer Singleton	Infinite Taxonomic Definite	High High Ignore High
Decision threshold	T242TDSH	Figure 6.1	Dempster-Shafer Singleton	Infinite Taxonomic High	High High Ignore High
Decision threshold	T242TDSL	Figure 6.1	Dempster-Shafer Singleton	Infinite Taxonomic Low	High High Ignore High
Decision threshold	T243SDSC	Figure 6.1	Dempster-Shafer Singleton	Infinite Sensor-state Definite	High High Ignore Good
Decision threshold	T243SDSH	Figure 6.1	Dempster-Shafer Singleton	Infinite Sensor-state High	High High Ignore Good
Decision threshold	T243SDSL	Figure 6.1	Dempster-Shafer Singleton	Infinite Sensor-state Low	High High Ignore Good

Test Description	TestID	Thesis Reference	Decision Algorithm/ Belief Representation	Conflict Threshold/ Ordering/ Decision Threshold	Abundant/ Established/ Rare/ Absent Data Quality
Conflict resolution	T257TCTH	Table 6.21	Dempster-Shafer Singleton	High Taxonomic Definite	High High Ignore Good
Conflict resolution	T257TCTM	Table 6.21	Dempster-Shafer Singleton	Medium Taxonomic Definite	High High Ignore Good
Conflict resolution	T258SCTH	Tables 6.19, 6.20	Dempster-Shafer Singleton	High Taxonomic Definite	High High Ignore Good
Conflict resolution	T258SCTM	Table 6.17	Dempster-Shafer Singleton	Medium Taxonomic Definite	High High Ignore Good
Indicator Value I = 5	T286_I05	Table 6.17	Dempster-Shafer Singleton	Infinite Sensor-state Definite	High High Ignore High
Indicator Value I = 10	T286_I10	Table 6.17	Dempster-Shafer Singleton	Infinite Sensor-state Definite	High High Ignore High
Indicator Value I = 20	T286_I20	Table 6.17	Dempster-Shafer Singleton	Infinite Sensor-state Definite	High High Ignore High
Indicator Value I = 30	T286_I30	Tables 6.17, 6.18	Dempster-Shafer Singleton	Infinite Sensor-state Definite	High High Ignore High
Simple support: Present Data	T051	Tables 7.1,7.2	Dempster-Shafer Simple support	Infinite Sensor-state Definite	High High Ignore Ignore
Simple-support:Present & Confirming Absence	T053	Tables 7.3, 7.4	Dempster-Shafer Simple support	Infinite Sensor-state Definite	High High Ignore Ignore
Simple-support:Present & Disconfirming Absence	T055	Tables 7.5, 7.6	Dempster-Shafer Simple support	Infinite Sensor-state Definite	High High Ignore High
Simple-support:Present & Disconfirming Absence (Variable weight)	T055Var	Tables 7.7, 7.8	Dempster-Shafer Simple support	Infinite Sensor-state Definite	High High Ignore Variable
Simple support. Present Data.	T057	Tables 7.9, 7.10, 7.11	Demspter-Shafer (T-R combination) Simple support	Infinite Sensor-state Definite	High High Ignore Ignore

Test Description	**TestID**	**Thesis Reference**	**Decision Algorithm/ Belief Representation**	**Conflict Threshold/ Ordering/ Decision Threshold**	**Abundant/ Established/ Rare/ Absent Data Quality**
Consonant belief. Present Data	T311	Table 7.12	Dempster-Shafer Consonant	Infinite Sensor-state Definite	High High Ignore Ignore
Consonant belief. Present & Absent Data	T313	Table 7.12	Dempster-Shafer Consonant	Infinite Sensor-state Definite	High High Ignore High

Glossary and Abbreviations

ϵ The degree of evidential discount used to represent the quality of a data source in uncertain reasoning.

κ The degree of evidential conflict in *Dempster's rule*.

σ In *Barnett's scheme*, a measure of uncertainty of a proposition.

Θ In *D-S* theory, the *environment* or *frame of discernment*.

a_i In *Barnett's scheme*, the evidence accumulated against (disconfirming) H_i.

AI Artificial Intelligence. The study of computational tasks that apparently require human intelligence.

ASPT Average Score per Taxon. Derived from the *BMWP* Score.

Autecology The *ecology* of a single species. Autecological studies are concerned with the correlation of the distribution of a species with environmental factors.

Automatic classifier A computer program incorporating an uncertain-reasoning decision algorithm, developed as part of the experimental work described in this thesis.

Barnett's scheme A computational scheme for combining evidence represented as *simple support* functions in linear time.

BBI Belgian Biotic Index. A European variant of the *TBI*.

Benthic organisms Organisms living on the bottom of a stream, lake, or the sea. Also referred to as the benthos.

BERT Benthic Ecology Response Translator. Name given to proposed expert or decision-support system for river water quality assessment using biological methods.

Bioassessment In the context of this thesis, the use of biological material to determine the quality of water.

Biocenosis A community occupying a given biotope.

Biota The living component of any system, e.g. of the hydrosphere, of an ecosystem.

BMWP Biological Monitoring Working Party, after which is named the BMWP Score.

BOD Biological Oxygen Demand. A measure of the amount of oxygen required for the aerobic micro-organisms to oxidise biochemically degradable aerobic matter in a water sample to a stable inorganic form.

bpa Basic probability assignment: the degree to which some evidence supports the various propositions within the *frame of discernment*.

c_i In *Barnett's scheme*, the measure of support against H_i.

CF Certainty Factor. Refers to the method of certainty factors, an uncertain-reasoning system which encodes measures of belief and disbelief.

Consonant belief A belief function assigned to a hierarchically-ordered set of focal elements within a *frame of discernment*.

D Shannon-Wiener (or Shannon-Weaver) diversity index.

D_i The degree of support for each hypothesis, as output by an *automatic classifier* as used in this project.

D-S Dempster-Shafer. Refers to Dempster-Shafer theory, considered to be a generalisation of the Bayesian calculus.

Dempster's rule The procedure for combining belief functions is *D-S* theory.

Determinand A general name for a characteristic or aspect of water quality; usually an attribute which can be numerically quantified.

DoE Department of the Environment.

DSS Decision Support System. A software system often incorporating graphics, databases and algorithmic models for solving broad or ill-defined problems in some domain.

e_i In *Barnett's scheme*, a simple evidence function formed from the combination of f_i and a_i.

EC European Community. Now the European Union.

ECI Expert Classification Index. Used in an earlier analysis for encoding the expert's intermediate classifications.

Ecology The study of the interrelation between living organisms and their environment.

Environment See *frame of discernment*.

Evidential discount See ϵ.

Expert system A computer system which emulates the decision-making ability of a human expert.

f_i In *Barnett's scheme*, the evidence accumulated for (confirming) H_i.

FBA Freshwater Biological Association.

Frame of Discernment The sample space of mutually exclusive and exhaustive hypotheses in Dempster-Shafer theory.

g_i Weighting factor for a species, as used in the calculation of Sládeček's Extended *Saprobic Index*.

GQA General Quality Assessment. Refers to a river quality assessment scheme recently proposed by the Department of the Environment.

H_i Refers to an hypothesis or several hypotheses in a sample space. $i = 1 \ldots n$, where n is the total number of singleton hypotheses.

h_i Abundance level of species i, as used in the calculation of Sládeček's Extended *Saprobic Index*.

IFE Institute of Freshwater Ecology. The IFE was established in April 1989 by agreement between the National Environment Research Council and the Freshwater Biological Association.

K_i In *Barnett's scheme*, the degree of internal evidential conflict for hypothesis H_i.

***I*-value** Indicator value of probability distribution, as devised for this project. For classifier output, the *I*-value encodes the "strength" of the distribution and the adjacency of the quality class rank-ordering.

ID3 Iterative Dichotomiser 3. Name given to a machine learning algorithm.

INTELLIPATH A commercial belief-network knowledge-based system for the diagnosis of lymph node diseases.

Invertebrates Animals not possessing a backbone.

K In *Barnett's scheme*, the degree of evidential conflict between hypotheses.

Knowledge-based system See *Expert System*.

L_s Likelihood of sufficiency, used in the odds-likelihood formulation of Bayes' rule.

L_n Likelihood of necessity. See L_s.

LEONARDO Name of software package for personal computers for developing *knowledge-based systems*.

LQI Lincoln Quality Index.

m In *D-S* theory, the basic probability assignment (*bpa*).

MSE Mean-square error. Referred to square of absolute difference between *ECI* and *SCI*. (Used in an earlier analysis).

MUNIN A belief-network system for diagnosis of neuromuscular disorders.

MYCIN An early celebrated expert system for diagnosing microbial infections.

NP Non-deterministic, Polynomial time. "NP-hard" refers to a class of non-linear computational problems.

NRA National Rivers Authority (now part of the Environment Agency).

NWC National Water Council.

O The decision order output by an *automatic classifier* as used in this project.

OASIS Name given to a graphical decision support system (*DSS*) for ground-water contaminant modelling.

Oxygen sag The de-oxygenation of river water caused by the discharge of organic wastes. Characteristic oxygen sag curves show that the point of maximum de-oxygenation usually occurs a considerable distance below the point of discharge. See also *BOD*.

p_i In *Barnett's scheme*, the measure of support for H_i.

Potable supply Supply of water suitable for human ingestion.

r_i In *Barnett's scheme*, the uncommitted belief (i.e. the support for Θ).

RAISON Acronym for "Regional Analysis by Intelligent Systems On a microcomputer". A knowledge-based system for modelling watershed acidification.

RIVPACS River Invertebrate Prediction and Classification System. A computer system for the classification of sites and the prediction of freshwater biological communities.

s In *D-S* theory, the degree of support assigned to an hypothesis.

s_i Preferred saprobic zone of a species *i*. Used in the calculation of Sládeček's Extended *Saprobic Index*.

S Saprobic Index.

SCI System Classification Index. Encodes the output of an *automatic classifier* in terms of an intermediate classification.

Simple support A belief function which assigns probability mass to singleton subsets or their complements.

Singleton support A term used in this thesis to represent an assignment scheme in which *evidential discount* is represented by a *bpa* to Θ, with the remainder assigned to singleton hypotheses.

SSSI Site of Special Scientific Interest.

T-R Interval-valued evidence combination, an alternative to Dempster's rule.

TBI Trent Biotic Index.

TWINSPAN Two-way Indicator Species Analysis. A computer program for arranging multivariate data in an ordered two-way table for classification of the individuals and attributes.

YWA Yorkshire Water Authority. Succeeded by the *NRA* (Northumbria and Yorkshire Region).

References

Ahmad, K. and Griffin, S. (1991) The acquisition of knowledge for controlling and rehabilitating network. In: Cedo Maksimovic (Ed.). *New Technologies in Urban Drainage.*

Allen, R.E. (Ed.) (1990) *The Concise Oxford Dictionary of current English.* Clarendon Press, Oxford.

AIAI (1990) *Course Manual for Knowledge Engineering.* Artificial Intelligence Applications Institute, University of Edinburgh, Scotland.

Andress, K.M. and Kak, A.C. (1987) A production system environment for integrating knowledge with vision data. Spatial Reasoning and Multi-sensor Fusion: Proceedings of the 1987 Workshop.

Armitage, P.D., Moss, D., Wright, J.F. and Furse, M.T. (1983) The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Research*, **17**, 333-47.

Armitage, P.D. (1989) The application of a classification and prediction technique based on macroinvertebrates to assess the effects of river regulation. In: James A. Gore and Geoffrey E. Petts (Eds.). *Alternatives in Regulated River Management.* CRC Press, Florida, U.S.A.

Barnett, J.A. (1981) Computational Methods for a Mathematical Theory of Evidence. Proceedings of the Seventh International Joint Conference on Artificial Intelligence.

Barron, J.J. (1993) Putting Fuzzy Logic into Focus. *Byte*, April, 111-118.

Bogler, P.L. (1987) Shafer-Dempster Reasoning with Applications to Multisensor Target Identification Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, Nov/Dec SMC-17, **6**, 968-977.

Bonissone, P.P. (1987) Plausible Reasoning: Coping with Uncertainty in Expert Systems. In: Stuart Shapiro (Ed.). *Encyclopaedia of Artificial Intelligence.* John Wiley & Sons Co., New York.

Bonissone, P.P. and Tong, R.M. (1985) Editorial: Reasoning with uncertainty in expert systems. *International Journal of Man-Machine Studies*, **22**, 241-250.

Boyd M., Walley, W.J. and Hawkes, H.A. (1993) Dempster-Shafer Reasoning for the Biological Surveillance of River Water Quality. Proceedings of the Second International Conference on Water Pollution. Computational Mechanics, Southampton.

Brooker, M.P. (1984) Biological Surveillance in Welsh rivers for Water quality and conservation assessment. In: D. Pascoe and R. W. Edwards (Eds.). *Freshwater Biological Monitoring*. International Association on Water Pollution Research and Control, Pergamon, Oxford.

Bryant, G.D. and Norman, G.R. (1980) Expressions of Probability: words and numbers. *New England Journal of Medicine*, 302, 7, 411.

Buchanan, B.G. and Shortliffe, E.H. (1985) *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project.* Addison-Wesley, Reading, Mass. London.

Budescu, D.V., Weinberg, S. and Wallsten, T.S. (1988) Decisions based on Numerically and Verbally Expressed Uncertainties. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 281-294.

Burton, A.M., Shadbolt, N.R., Rugg, G. and Hedgecock, A.P. (1990) The efficacy of knowledge elicitation techniques: a comparison across domains and levels of expertise. *Knowledge Acquisition,* **2**, 167-178.

Caselton, W.F., Froese, T.M., Russell, A.D. and Luo, W. (1988) Belief input procedures for Dempster-Shafer based expert systems. In: J. S. Gero (Ed.). *Artificial Intelligence in Engineering: Robotics and Processes.* Elsevier: Computational Mechanics Publications, Southampton.

Chandler, J.R. (1970) A Biological Approach to Water Quality Management. *Water Pollution Control,* **69**, 415-422.

Chang, L.W. and Kashyap, R.L. (1990) Study of Interval-Valued Belief Combinations with Conflicting Evidence. Proceedings of the 2nd International IEEE Conference on Tools for Artificial Intelligence. IEEE, Piscataway, NJ, USA.

Cheng, Y.Z. and Kashyap, R.L. (1988) An Axiomatic Approach for Combining Evidence from a Variety of Sources. *Journal of Intelligent and Robotic Systems*, **1**, 17-33.

Chapman, D. (1992) *Water quality assessments: a guide to the use of biota, sediment and water in environmental management.* Chapman & Hall on behalf of UNESCO, WHO and UNEP, London.

Charniak, E. (1991) Bayesian Networks without Tears. *AI Magazine*, Winter, 50-63.

Cheeseman, P. (1985) In Defense of Probability. Proceedings of the Ninth International Joint Conference on Artificial Intelligence.

Chesters, R.K. (1980) Biological Monitoring Working Party. The 1978 national testing exercise. Department of the Environment. Water Data Unit, Technical Memorandum 19.

Cohen, P.R. (1985) *Heuristic Reasoning About Uncertainty: An Artificial Intelligence Approach.* Pitman, Boston, MA, U.S.A.

Collins, K. (1993) The Political Perspective. River Water Quality Monitoring and Control. Proceedings of the Freshwater Europe Symposium. Aston University, Birmingham.

Cortes-Rello, E. and Golshani, F. (1990) Uncertain reasoning using the Dempster-Shafer method: an application in forecasting and marketing management. *Expert Systems*, 7, February No. 1, 9-17.

Creative Logic (1992) LEONARDO User's Guide. Creative Logic, London.

Cullen, J. and Bryman, A. (1988) The knowledge acquisition bottleneck: time for reassessment? *Expert Systems*, **5**, 216-225.

De Pauw, N. and Hawkes, H.A. (1993) Biological Monitoring of River Water Quality. River Water Quality Monitoring and Control: Proceedings of the Freshwater Europe Symposium. Aston University, Birmingham.

Dempster, A.P. (1967) Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, **38**, 325-339.

Department of the Environment (1972) Report of a River Pollution Survey of England and Wales, 1970. HMSO, London.

Department of the Environment (1980) Assessment and presentation of the biological qualities of rivers in Great Britain. Final report of the Biological Monitoring Working Party of the Standing Technical Advisory Committee on Water Quality.

Department of the Environment, Welsh Office (1992) River Quality, The Governments's Proposals: A Consultation Paper. Department of the Environment, Great Britain.

Department of the Environment Welsh Office (1993) The Surface Waters (Fisheries Ecosystem) (Classification) Regulations 1993: Consultation Paper. Department of the Environment, Great Britain.

Dillard, R.A. (1992) Using Data-Quality Measures in Decision-Making Algorithms. *IEEE Expert*, 63-72.

Dong, W.-M. and Wong, F.S. (1986a) From uncertainty to approximate reasoning: part 1: conceptual models and engineering interpretations. *Civil Engineering Systems*, **3**, September, 143-154.

Dong, W.-M. and Wong, F.S. (1986b) From uncertainty to approximate reasoning: part 2: reasoning with algorithmic rules. *Civil Engineering Systems*, **3**, December, 192-202.

Dubois, D. and Prade, H. (1985) Combination and Propagation of Uncertainty with Belief Functions. Proceedings of the Ninth International Conference on Artificial Intelligence.

Duda, R.O., Hart, P.E. and Nilsson, N.J. (1976) Subjective Bayesian Methods for Rule-based Inference Systems. Proceedings of the National Computer Conference.

Extence, C.A., Bates, A.J., Forbes, W.J. and Barham, P.J. (1987) Biologically Based Water Quality Management. *Environmental Pollution*, **45**, 221-236.

Friedrich, G. (1990) Eine Revision des Saprobiensystems. *Zeitschrift für Wasser und Abwasserforschung*, **23**, 141-152.

Friedrich, G., Chapman, D. and Beim, A. (1992) The Use of Biological Material. In: D. Chapman (Ed.). *Water quality assessments: a guide to the use of biota, sediments and water in environmental monitoring*. Chapman & Hall on behalf of UNESCO, WHO and UNEP, London.

Garvey, T.D., Lowrance, J.D. and Fischler, M.A. (1981) An Inference Technique for Integrating Knowledge from Disparate Sources. Proceedings of the Seventh International Joint Conference on Artificial Intelligence.

Genest, C. and Zidek, J.V. (1986) Combining Probability Distributions: A Critique and Annotated Bibliography. *Statistical Science*, **1**,114-148.

Gero, J.S. (Ed.) (1988) *Artificial Intelligence in Engineering: robotics and processes*. Elsevier (with Computational Mechanics Publications)., Southampton.

Giarratano, J. and Riley, G. (1989) *Expert Systems: Principles and Programming*. PWS-Kent, Boston.

Gordon, J. and Shortliffe, E.H. (1985a) The Dempster-Shafer Theory of Evidence. In: Bruce G. Buchanan and Edward H. Shortliffe (Eds.). *Rule-Based Expert Systems: the MYCIN experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading, Mass. London.

Gordon, J. and Shortliffe, E.H. (1985b) A Method for Managing Evidential Reasoning in a Hierarchical Hypothesis Space. *Artificial Intelligence*, **26**, 323-357.

Graham, I. and Jones, P.L. (1988) *Expert systems: knowledge, uncertainty and decision*. Chapman and Hall, London.

Green, R.H. (1979) *Sampling Design and Statistical Analysis for environmental biologists*. Wiley, New York, U.S.A.

Grzymala-Busse, J.W. (1991) *Managing Uncertainty in Expert Systems*. Kluwer Academic Publishers, Massachusetts, U.S.A.

Hart, A. (1989) *Knowledge acquisition for Expert Systems*. Kogan Page, London.

Hau, H.L. (1990) Decomposition of belief function in hierarchical hypothesis space. Proceedings of the 2nd International IEEE Conference on Tools for Artificial Intelligence. IEEE, Piscataway, NJ, USA.

Hawkes, H.A. (1956) The biological assessment of pollution in Birmingham streams. *J. Inst. Munic. Engrs.*, **82**, 452-63.

Hawkes, H.A. (1964) An Ecological Basis for the Biological Assessment of River Pollution. *Chem. Ind.*, 437-442.

Hawkes, H.A. (1974) Water quality: biological considerations. *Chemistry and Industry*, 21 December, 990-1000.

Hawkes, H.A. (1975) River zonation and classification. In: B. A. Whitton (Ed.). *River Ecology*. Blackwell, Oxford.

Hawkes, H.A. (1977) *Manual on Biological Surveillance of Rivers using Benthic Macroinvertebrates*. Aston University, Birmingham.

Hawkes, H.A. (1978) Conceptual basis for the biological surveillance of river water quality. In: J. G. Hughes and H. A. Hawkes (Eds.). Biological surveillance of river water quality. Proceedings of Section K, jointly with Section D, of the British Association for the Advancement of Science, Aston 1977., University of Aston.

Hawkes, H.A. (1979a) Invertebrates as indicators of river water quality. In: A. Jones and L. Evison (Eds.). *Biological indicators of water quality*. John Wiley & Sons, Chichester.

Hawkes, H.A. (1979b) Water quality issues: an ecological reaction. *Chemistry and Industry*, 17 March, 201-204.

Hawkes, H.A. (1981) Biological surveillance of rivers in relation to fisheries - potential and limitations. Proceedings of the 2nd. British Freshwater Fisheries Conference.

Hawkes, H.A. and Davies, L.J. (1971) Some effects of organic enrichment on benthic invertebrate communities in stream riffles. The 11th. Symposium of The British Ecological Society. Blackwell Scientific Publications, Oxford.

Hellawell, J.M. (1977) Biological surveillance and water quality monitoring. In: J. S. Alabaster (Ed.) *Biological monitoring of inland fisheries*. Applied Science Publishers Ltd., London.

Hellawell, J.M. (1978) *Biological surveillance of rivers*. Water Research Centre, Medmenham.

Hellawell, J.M. (1986) *Biological Indicators of Freshwater Pollution and Environmental Management*. Elsevier Applied Sciences Publishers, London.

Henrion, M. (1990) An Introduction to Algorithms for Inference in Belief Networks. In: R. Oliver and J. Smith (Eds.). *Influence Diagrams, Belief Nets, and Decision Analysis*. Wiley, Chichester.

Henrion, M. (1988) Propagation of Uncertainty by Probabilistic Logic Sampling. In: T. Levitt, J. Lemmer and L. Kanal (Eds.). *Uncertainty in Artificial Intelligence 3*. North Holland, Amsterdam.

Henrion, M., Breese, J.S. and Horvitz, E.J. (1991) Decision Analysis and Expert Systems. *AI Magazine*, Winter, 65-90.

Hill, M.O. (1979) *TWINSPAN - A FORTRAN Program for Arranging Multivariate Data* *in an Ordered Two-Way Table by Classification of the Individuals and Attributes*. Ecology and Systematics, Cornell University, Ithaca, New York.

Horvitz, E.J., Breese, J.S. and Henrion, M. (1988) Decision Theory in Expert Systems and Artificial Intelligence. *International Journal of Approximate Reasoning - Special Issue on Uncertainty and Artificial Intelligence*, **2**, 247-302.

Hynes, H.B.N. (1960) *The Biology of Polluted Waters*. Liverpool University Press, Liverpool.

Hynes, H.B.N. (1970) *The Ecology of Running Waters*. Liverpool University Press, Liverpool.

Illies, J. (Ed.) (1967) *Limnofauna Europaea*. Gustav Fischer Verlag, Stuttgart and New York, U.S.A.

Ip, H.H.S and Wong, H.-M. (1991) Evidential reasoning in foreign exchange rates forecasting. The First International Conference on Artificial Intelligence on Wall Street (Cat. No. 91TH0399-6). IEEE Computer Society Press, Los Alamitos, CA, USA.

Jaccard, P. (1912) The distribution of the flora in the alpine zone. *New Phytologist*, **11**, 37-50.

Jeffries, M. and Mills, D. (1990) *Freshwater Ecology: principles and applications*. Belhaven Press, London.

Johnson, L. and Keravnou, E.T. (1985) *Expert Systems Technology: a guide*. Abacus Press.

Kolkwitz, R. and Marsson, M. (1908) Ökologie der planzlichen Saprobien. (Ecology of the plant saprobien). *Ber. Dt. Botan. Ges.*,**261**, 505-19.

Kolkwitz, R. and Marsson, M. (1909) Ökologie der tierischen Saprobien. (Ecology of the animal saprobien). *Int. Rev. ges. Hydrobiol.*,**2**, 125-52.

Kyburg, H.E., Jr. (1987) Bayesian and Non-Bayesian Evidential Updating. *Artificial Intelligence*, **31**, 3, 271-293.

Lam, D.C.L., Swayne, D.A., Fraser, A.S., Storey, J. and Wong, I. (1989a) Expert systems in modelling of regional water quality. *Modelling and Simulation*, **20**, 2, 651-658.

Lam, D.C.L., Swayne, D.A., Storey, J. and Fraser, A.S. (1989b) Watershed acidification using the knowledge-based approach. *Ecological Modelling*, **47**, 131-152.

Lam, D.C.L., Wong, I., Swayne, D.A., Storey, J. and Kerby, J.P. (1990) Application of the Raison Expert System for Water Pollution Problems: From Acid Rain to Mine Effluent. In: P. Zannatti (Ed.). *Computer Techniques in Environmental Studies III*. Springer-Verlag, Germany.

Lindley, D.V. (1987) The probability approach to the treatment of uncertainty in artificial intelligence and expert systems. *Statistical Science*, **2**, 1, 17-24.

Lingras, P. and Wong, S.K.M. (1990) Two Perspectives of the Dempster-Shafer theory of belief functions. *International Journal of Man-Machine Studies*, **33**, 467-487.

Maitland, P.S. (1977) *A Coded Checklist of Animals occurring in Fresh Water in the British Isles*. Institute of Terrestrial Ecology, Edinburgh.

Martin, J.R. (1993) Costs and Benefits of Chemical Monitoring. River Water Quality Monitoring and Control: Proceedings of the Freshwater Europe Symposium. Aston University, Birmingham.

Martin-Clouaire, R. and Prade, H. (1985) On the problems of representation and propagation of uncertainty in expert systems. *International Journal of Man-Machine Studies*, **22**, 251-264.

Mason, C.F. (1991) *Biology of Freshwater Pollution*. Longman Scientific and Technical, London.

McLelland, J.L. and Rummelhart, D.E. (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* MIT Press, Cambridge, MA, U.S.A.

Metcalfe, J.L. (1989) Biological Water Quality Assessment of Running Waters Based on Macroinvertebrate Communities: History and Present Status in Europe. *Environmental Pollution*, 60, 101-139.

Meybeck, M., Friedrich, G., Thomas R. and Chapman, D. (1992) Rivers. In: D. Chapman (Ed.) *Water quality assessments: a guide to the use of biota, sediments and water in environmental monitoring*. Chapman and Hall on behalf of UNESCO, WHO and UNEP, London.

Michalski, R. and Chilausky, R. (1980) Learning by being told and learning from examples. *International Journal of Man-Machine Studies*, **12**, 63-87.

Michie, D. (1979) *Expert Systems in the Microelectronic Age*. Edinburgh University Press, Edinburgh.

Murphy, P.M. (1978) The temporal variability in biotic indices. *Environmental Pollution*, **17**, 227-36.

Nakao, M.A. and Axelrod, S. (1988) Numbers are better than words. *American Journal of Medicine*, **74**, 1061-1065.

National Rivers Authority (1991a) *The Quality of Rivers, Canals and Estuaries in England and Wales*. Report of the 1990 Survey. Water Quality Series No. 4. December 1991. National Rivers Authority, Rivers House, Waterside Drive, Aztec West, Almondsbury, Bristol.

National Rivers Authority (1991b) *Proposals for Statutory Water Quality Objectives*. Water

Quality Series No. 5. December 1991. National Rivers Authority, Rivers House, Waterside Drive, Aztec West, Almondsbury, Bristol.

Neapolitan, R.E. (1990) *Probabilistic Reasoning in Expert Systems*. John Wiley & Sons, New York.

Ng, K. and Abramson, B. (1990) Uncertainty management in expert systems. *IEEE Expert*, **5**, 2, 29-48.

Open University (1989) *Knowledge Engineering*. Open University, Milton Keynes.

Pantle, R. and Buck. H. (1955) Die biologische Überwaschung der Gewässer und die Darstellung der Ergebrisse. (Biological monitoring of water bodies and the presentation of results). *Gas und Wasserfach*, **96**, 104.

Pearl, J. (1982) Reverend Bayes on inference engines: A distributed hierarchical approach. In: *Proceedings Second National Conference on Artificial Intelligence*, Pittsburgh, PA, 133-136.

Pearl, J. (1986a) Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, **29**, 241-288.

Pearl, J. (1986b) On Evidential reasoning in a hierarchy of hypotheses. *Artificial Intelligence*, **28**, 9-15.

Pearl, J. (1987) Bayesian decision methods. In: Stuart Charles Shapiro (Ed.). *Encyclopaedia of AI.* Wiley Interscience, New York, U.S.A.

Pearl, J. (1988) *Probabilistic reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA.,U.S.A.

Pendle, D.R. and Franklin, S.E. (1992) Multisource evidential classification of surface cover and frozen ground. *International Journal of Remote Sensing*, **13**, Number 17 Nov 20, 3375-3380.

Persoone, G. and De Pauw, N. (1979) Systems of biological indicators for water quality assessment. In: O. Ravera (Ed.). Biological aspects of freshwater pollution: proceedings of the course held at the Joint Research Centre of the Commission of the European Communities, Ispra, Italy, 5-9 June 1978. Pergamon for the Commission of the European Communities, Oxford.

Principe, J.C., Gala, S.K. and Chang, T.G. (1989) Sleep Staging Automaton Based on the Theory of Evidence. *IEEE Transactions on Biomedical Engineering*, **36**, 5, 503-509.

Ruck, B.M. (1995) *The Application of Artificial Neural Networks to the Interpretation and Classification of Freshwater Benthic Invertebrate Communities*. PhD thesis, Department of Civil Engineering, University of Aston, Birmingham, England.

Shachter, R.D. (1986) Evaluating Influence Diagrams. *Operations Research*, **34**, 871-882.

Shachter, R.D. and Heckerman, D. (1987) Thinking Backward for Knowledge Acquisition. *Artificial Intelligence*, **8**, 55-63.

Severn-Trent Water Authority (1988) *Water Quality 1987/88. Appendix 7 River Quality - Biological and Fisheries*. Severn-Trent Water Authority, Birmingham.

Shadbolt, N. (1989) Knowledge representation in man and machine. In: R. Forsyth (Ed.). *Expert Systems: Principles and Case Studies*. Chapman and Hall, London.

Shafer, G. and Logan, R. (1987) Implementing Dempster's rule for hierarchical evidence. *Artificial Intelligence*, **33**, 271-298.

Shafer, G. (1976) *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, U.S.A.

Shafer, G. (1987) Probability Judgement in Artificial Intelligence and Expert Systems. *Statistical Science*, **2**, 1, 3-44.

Shafer, G. and Pearl, J. (Eds.) (1990). *Readings in Uncertain Reasoning*. Morgan Kaufman, California, USA.

Shortliffe, E.H. and Buchanan, B.G. (1985) A Model of Inexact Reasoning In Medicine. In: Bruce G. Buchanan and Edward H. Shortliffe (Eds.). *Rule-Based Expert Systems: the MYCIN experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading, Mass. London.

Sládeček, V. (1973) System of water quality from the biological point of view. *Arch. Hydrobiol. Beih. Ergebn, Limnol.*, **7**(I-IV), 1-218.

Sucar, L.E., Gillies, D.F. and Gillies, D.A. (1993) Objective probabilities in expert systems. *Artificial Intelligence,* **61**, 187-208.

Sweeting, R., Lowson, D., Hale, P. and Wright, J. (1992) 1990 Biological Assessment of Rivers in the UK. In: P. J. Newman, M.A. Piavaux and R.A. Sweeting (Eds.). *River Water Quality - Ecological assessment and Control*. Commission of the European Communities, Brussels.

Szolovits, P. and Pauker, S. (1978) Categorical and probabilistic reasoning in medical diagnosis. *Artificial Intelligence*, **11**, 115-144.

Tanimoto, S. L. (1990) *The elements of artificial intelligence*. Computer Science Press, New York, U.S.A.

Tonn, B.E. and Goeltz, R. (1990) Psychological validity of uncertainty combining rules in expert systems. *Expert Systems*, **7**, 2, May, 94-101.

Tonn, B.E. and Goeltz, R. (1992) Eliciting reliable uncertainty estimates. *Expert Systems*, **9**, 1, February, 25-33.

Tversky, A. and Kahneman, D. (1974) Judgement under uncertainty: heuristics and biases. *Science*, **185**, 1124-1131.

Voorbraak, F. (1990) A computationally efficient approximation of Dempster-Shafer theory. In: B.R. Gaines and J.H. Boose (Eds.), Machine Learning and Uncertain Reasoning. Academic Press, London.

Voorbraak, F. (1991) On the justification of Dempster's rule of combination. *Artificial Intelligence*, **48**, 171-197.

Walley, W.J., Hawkes, H.A. and Boyd M. (1992a) Application of Bayesian Inference to River Water Quality Surveillance. Applications of Artificial Intelligence in Engineering VII. Elsevier: Computational Mechanics Publications, Southampton.

Walley, W.J., Boyd M. and Hawkes, H.A. (1992b) An expert system for the biological monitoring of river pollution. Proceedings of the Fourth International Conference on Computer Techniques in Environmental Studies. Elsevier, Portsmouth, England.

Warren, C.E. and Doudoroff, P. (1971) *Biology and Water Pollution Control.* W.B. Saunders Company, London.

Washington, H.G. (1984) Diversity, biotic and similarity indices. A review with special relevance to aquatic ecosystems. *Water Research*, **18**, 653.

Wilhm, J.L. and Dorris, T.C. (1968) Biological parameters for water quality criteria. *Bioscience*, Part 18, 477-481.

Whitton, B.A. (Ed.) (1975) *River Ecology.* Blackwell Scientific Publications, Oxford.

Wong, F., Dong, W., Boissonnade, A. and Ross, T. (1986) Expert opinions and expert systems. Proceedings of the 9th ASCE Conference on Electronic Computing. American Society of Civil Engineers, New York, U.S.A.

Woodiwiss, F.S. (1964) The biological system of stream classification used by the Trent River Board. *Chem. Ind.*, 443-447.

Wright, J.F., Armitage, P.D., Furse, M.T. and Moss, D. (1988) A new approach to the biological surveillance of river quality using macroinvertebrates. *Verh. Internat. Verein. Limnol.*, **23**, October, 1548-1552.

Wright, J. F., Armitage, P. D., Furse, M. T. and Moss, D. (1989) Prediction of invertebrate communities using stream measurements. *Regulated Rivers: Research and Management*, **4**, 147-155.

Yorkshire Water Authority (1976) The Collection of Invertebrate Distribution Records 1971-76. Scientific Services, Yorkshire Water Authority, Leeds.

Zadeh, L.A. (1965) Fuzzy Sets. *Information and Control*, **8**, 338-353.

Zadeh, L.A. (1978) Fuzzy Sets as a Basis for a Theory of Possibility. *Fuzzy Sets and Systems*, **1**, 3-28.

Zadeh, L.A. (1984) Review of a mathematical theory of evidence, by G. Shafer. *AI Magazine,* 81-83.

Zadeh, L.A. (1986) A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. *The AI Magazine*, Summer, 85-90.

Zadeh, L A. (1988) Fuzzy Logic. *IEEE Computer*, April, 83-93.

Zelinka, M. and Marvan, P. (1961) Zur Präzisierung der biologischen Klassifikation der Reinheit fliessender Gewässer. (Making a precise biological classification of the quality of running waters). *Arch. Hydrobiol.*, **57**, 389-407.