4 BugStats: Software for Environmental Reconstruction and Statistics from Beetle Assemblages

BugStats is the statistics component of the BugsCEP system, and includes a number of (semi-) quantitative tools to assist in the interpretation of fossil, and modern, beetle assemblages. The primary function of BugStats is the production of easily read habitat diagrams from countsheets, which can assist in the interpretation of the environmental implications of faunas, collectively referred to as 'EcoFigs' in the software and this thesis. The majority of taxa in BugsCEP have been ascribed to 22 habitat groups, referred to as Bugs EcoCodes, and by summing these habitats, as represented by the taxa, it is possible to produce a summary of the environment represented by all the taxa in the sample. This can be undertaken for a sequence of samples, and output graphically in MS Excel format. The diagrams can then be interpreted with respect to insect ecology, biodiversity and taphonomy towards environmental reconstruction, be it a palaeo- or modern environment. Due to the nature of its reference data, BugStats does not produce *complete* environmental reconstructions, but rather reconstructions of those parts of the environment that can be represented by the beetle fauna. This is taken as implicit in the rest of this chapter, and it is acknowledged that the use of additional proxy data sources will almost always be able to complement the reconstructions created from beetle assemblages. There are also a number of taphonomic issues, which are discussed below (section 4.2.1.2), that one should be aware of when using data from any proxy data source, and these should be considered when using BugStats.

Palaeoentomology is essentially a science of analogy. That is to say that modern ecological information is projected onto the fossil assemblages with the assumption that the species have the same habitat requirements now as they did at the time of death. Archaeologists may find it useful to think of anthropological analogues, where modern ethnographic studies are used to infer details of prehistoric peoples. The principal is in essence the same, although the smaller number of variables involved, and the absence of 'the human factor', makes palaeoecological analogy potentially easier and more reliable.

BugStats is an aid to, and not a substitute for careful consideration of the ecological implications of the taxa found within samples and the environments represented by them. It includes tools for manipulating the results in order to compensate for differences in sample abundances or species numbers, which can be turned on or off at the discretion of the user. It should be remembered that even if variations in numbers hinder certain aspects of inter-sample comparison, these factors are real products of either the palaeoenvironment, depositional environment, taphonomic processes or sampling strategy, and should never be ignored or obscured in statistics when undertaking palaeoenvironmental reconstruction.

4.1 Why BugStats?

As mentioned in Chapter 1, palaeoentomology has not seen the same degree of quantification and software development as a number of other proxy based fields. Although this in itself is reason enough to push forward the creation of new methods and tools, there are several principle reasons for the development of BugStats within the BugsCEP program.

1. There is a need for a standardized (i.e. regionally comparable), visual and transparent system for summarising quantitative palaeoentomological interpretations and reconstructions. This should be able to work for faunas with very small numbers of individuals or taxa, as well as more complex assemblages. The system should be transparent enough that a non-expert[i] can understand it.

---

[i] I.e. non-palaeoentomologist

2. There is a need to compliment existing statistical tools with some more specifically applicable to palaeoentomology. Ordination methods (PCA, CA, etc.[ii]) and cluster analyses, which have been used to examine fossil beetle assemblages earlier (e.g. Cong & Ashworth, 1997), do not easily produce results which allow inter-site comparisons, as they are dimensioned only according to the single site faunas. In addition, palaeo-data often violate or confuse assumptions necessary for many statistics, e.g. by way of taphonomic issues, many zeros, and irregular population structures in the data. The datasets involved are also often too small to provide traditionally statistically robust results.

3. There is currently no existing software specifically designed to work with fossil insect data. The lack of (semi-)quantitative environmental reconstruction software makes it harder for palaeoentomologists to prepare results rapidly that have an immediate visual impact on non-experts.

4. There is a need for a system that is able to handle fossil and modern data on an equal footing, and thus allow for long term (palaeo)biodiversity studies.

4.2 Background

4.2.1 Quantitative methods in Quaternary science

Ecologists often talk in terms of response (dependent) variables, and explanatory (independent) variables (e.g. Jongman *et al.*, 1995) when designing and undertaking experiments. For example, in a study of the effect of soil moisture on modern carabid populations the explanatory variable might be soil moisture, and the response variable some measure of the carabid population (e.g. Luff *et al.*, 1989). Both of these variable types are usually measurable in the field or experimentally, and it is the task of the ecologist to establish the relationships between them. In palaeoecology on the other hand, only the response variable, i.e. the numbers of species and individuals, is measurable, and this will almost certainly have been filtered by taphonomic processes. It is then the task of the palaeoecologist to translate the response data into a measure of the explanatory variable using the available knowledge of the relationships between them. It follows from this reasoning that palaeoecology is at the mercy of ecology for the knowledge of the processes that control observable populations and variables, the latter being only observable in the past as translated by the observed fossil data. This is what is widely known as proxy analysis.

In palaeoenvironmental reconstruction the primary goal is some form of numerical or verbal representation of a past environment. As any particular proxy source can only help reconstruct a limited number of variables, it follows that the reconstructions created from any proxy must be expressed in the terms that the particular proxy allows. It also follows, that since different proxies respond to different sets of variables, the more proxies used the wider or more detailed the range of environments that can be reconstructed. This thesis, and this chapter in particular, only deals with beetles, however, and their part in the multi-proxy science of palaeoenvironmental reconstruction.

4.2.1.1 Statistics in (palaeo)entomology and environmental science

The statistical techniques applied to insect assemblages can be divided into two categories: descriptive and investigative/explanatory, although the boundary between these can, by use, be somewhat vague. Although it would seem logical that descriptive statistics are desirable if we are to reconstruct past environments, this is not entirely true. The fact that we are usually dealing with a set of unknown environmental variables means that we must first perform some form of investigative analysis in order to assess the variables that can be described. The development of the MCR method is a good example. *Investigative* (PCA) analyses by Atkinson *et al.* (1986) showed a clear link between temperature and the geographical distribution of certain species of Coleoptera. From this assumption a calibration

---

[ii] Principle Component Analysis, Correspondence Analysis.

dataset was produced which allowed fossil beetle assemblages to be used to *describe* palaeotemperatures (see Chapter 5 for more details).

In ecology the boundaries are often more clear cut, at least where the explanatory variables are measurable. This allows the researcher to construct models of population dynamics in response to changes in these variables. There are cases, however, such as when investigating the unknown causes of biodiversity changes or populations in poorly researched environments, where the methods employed may be similar to those in palaeoecology.

4.2.1.2 *Taphonomy and the representation of taxa in samples*

Although the level of identification of fossil specimens could be a function of the proficiency of the investigator, most palaeoentomologists will fully pursue identification of all reasonably whole fragments in any scientific investigation, although this is not always the case in consultancy work (cf. Kenward, 1992). Levels of identification could also reflect the ease of identification of a genus, the resilience of the sclerites[iii], and the potential for preservation of a specific group. The last is certainly true for the extreme end of the size range, for example, large diving beetles such as *Dytiscus marginalis* L. (27-35 mm[iv]) are usually highly fragmented in processing, and possibly deposition, if found fossil at all (BugsCEP has only six records, representing twelve individuals, for this species). At the other end of the size spectrum, tiny beetles such as *Latridius minutus* (grp.) (L.) (1-2 mm) are a common find in archaeological deposits, and may stand a greater chance of surviving as fossils due to their size (BugsCEP has c. 438 records, representing 2 719 individuals). On the other hand, members of the Pselaphidae family (moss beetles) which are c. 1 mm long are more rarely found fossil, and may easily be lost during processing, especially if a sieve coarser than 300 μm is used (BugsCEP has c. 891 records, representing 2 267 individuals over the 22 taxa within the family, equating to approximately 40 records and 103 individuals per taxon). There are also differences in the thickness of chitin and general form that may aid preservation. The weevils *Otiorhynchus nodosus* (Müll.) and *O. arcticus* (O. Fabricius) are frequent finds from relatively cold climate samples, and their round, robust abdomens survive when no other taxa are preserved. Other families, such as the Cantharidae (soldier beetles) and some of the Chrysomelidae (leaf beetles) have much more delicate sclerites and are more susceptible to erosion and fragmentation.

Whilst standardization or normalization of datasets is commonplace in many branches of ecology (see Jongman *et al.*, 1995), there are complications when applying the techniques to fossil assemblages. The primary concern is that the total abundance or species richness of a sample actually may mean something important in terms of the habitat represented by the fauna. For example, environments with more abundant insects would be expected to leave more fossils. Experience tells us, for example, that a Viking Age farm floor was much more beetle rich than an equivalent area of peat bog in the same region (Buckland *et al.,* 1993). Changes in sample abundances may also reflect changes in sedimentation rates, rather than environmental change. In addition, differential preservation, as discussed above, may also be represented in the totals. As long as these factors are taken into consideration, and a certain degree of approximation is accepted in the results, standardization can be used to balance the differential sample sizes within a sequence, and allow inter-sample comparison on an equal footing. The changes in habitat proportions represented in standardized diagrams should be less dependent on the total sample abundance or richness, and thus give a more reliable picture of environmental change. The methods available in BugStats are described in section 4.3.3.3.

4.2.2 Biogeography, evolution and palaeoecology

Although there is some disagreement between palaeoecologists and molecular biologists (among others), it is generally assumed that beetles have evolved very little over at least the past few million years (Coope, 1978). Some species have even been shown to exhibit morphological constancy over the

---

[iii] A sclerite is a piece of the beetle exoskeleton.
[iv] Size data from Harde, 1992.

past 30 million years (Elias, 1994). This is largely due to their ability to migrate as species during times of rapid climate change, and their early appearance in the evolutionary history of the planet – the earliest six legged arthropods being found in deposits over 400 million years old (Grimaldi & Engel, 2005). The concept of *species constancy* which derives from these findings, allows us to assume that species that exist at the present day occupied the same ecological niche in the past (Coope, 1978). Although rapid evolutionary changes at the molecular level have been identified by some authors (e.g. Clarke *et al.*, 2001), there is sufficient evidence to believe that we can rely on fossil beetles to be able to inform us on palaeoenvironments by analogy with their modern habitats. Revisions in taxonomy as a result of molecular studies rarely lead to changes that affect more than the names of taxa used in palaeoecology, which rely entirely on morphology for identification. Doubt over this principle occasionally occurs in the literature, and for further discussion see the introduction to Ashworth *et al.* (1997), and the references contained therein.

Particular care must be taken, however, where there is reason to believe that a population may have been restricted in its migration possibilities and survived as an isolated group. Such isolation, perhaps caused by an island or isolated mountain existence, may have led to allopatric speciation in populations of what were initially the same species. This debate is particularly active regarding the existence of glacial refugia, and is as much active in terms of plants as insects (see e.g. the debate between Tzedakis *et al.,* 2002; Stewart, 2003; Tzedakis *et al.*, 2003). Speciation among water beetles has been especially well studied (e.g. Drotz, 2003), but has had little impact on palaeoecology due to the low degree of differentiation in aquatic environments currently used in reconstructions. Such revisions may be more significant where the species are used in thermal reconstructions (see Chapter 5), for example, as could be the case with the suggestion that *Agabus solieri* Aubé is to be found at higher altitudes than the previously synonymous *Agabus bipustulatus* (L.) (Drotz, 2003). The two are very difficult to distinguish between on the fossil parts. Even when aware of such potential problems, it is not always easy in Quaternary science to decide what can be assumed about a palaeo-population, and it is therefore advisable to build a degree of flexibility into reconstructions, even where the quantification of errors is impossible.

Taxonomic classification does not always follow ecological classification, and there is significant variability in the range of habitats occupied by species at the family and even generic level. In the Bugs EcoCode classification system, the majority of water beetles of the family Hydrophilidae, for example, are classed as *Aquatics*, (although a number of them may inhabit dung/foul habitats), whereas ground beetles (Carabidae) are found in a large variety of terrestrial habitats. This is partly a result of the choice of classifiers, and the Bugs EcoCode system has a large terrestrial bias – it does not differentiate between more than running and standing water habitats, whereas its does differentiate between a wide variety of terrestrial habitats. This is a direct reflection of the purpose of the system, and the wide variety of aquatic habitats are not often of interest in palaeoecology at this coarse level of description, although the work on Caddis (Trichoptera) by Greenwood *et al.* (2006) illustrates the possibilities. A classification system specifically designed for aquatic environments would be quite different to the present Bugs EcoCodes, and may differentiate between vegetation regimes, depth and water quality among other things. Such a system would benefit greatly from the inclusion of orders beyond the Coleoptera, and is a target area for future improvements in BugsCEP.

4.2.3 Classification

The value of beetles as environmental indicators is well established (e.g. Ashworth *et al.*, 1997). The large numbers of species, and the variety of environments in which they are found, necessitate some form of summation, and the classification of species after their habitat requirements is a common choice. This not only provides a method of efficiently describing modern environmental change, but is also of significant benefit in palaeoenvironmental reconstruction and archaeoentomological interpretation (e.g. Robinson, 2001).

Ecological classification systems are often designed with specific aims, either within restricted areas of research, or geographically limited areas. For example, Davies *et al.,* (2002) "…classify the

Richards Bay [South Africa] dung beetle community into five groups based on both biogeographical distribution pattern and local vegetation association." with the aim of assessing the progress of habitat restoration in a landscape fragmented by mining. The important distinction should be noted between the classification of habitats by the species that occupy them – which are called *species groups* in this thesis, but 'habitat groups' by some authors – and the description of species that occupy habitats classified by other means (e.g. general description, vegetation survey and hydrology). The latter habitat groups tend to be more qualitative descriptions than quantitative definitions, and thus are more flexible and applicable to multi-region studies. The two terms do not necessarily describe the same habitats. For example, Eyre & Luff (1990) classify grassland habitats throughout Europe using the TWINSPAN software (Hill, 1979) to analyse the Carabid species found at 638 sites. They thus produce seventeen habitat groups which, although statistically valid, may be of limited applicability in some areas of Europe due to species assemblages which may not have equivalents in the original dataset. The range of habitat groups is limited to those available in the original dataset. BugStats uses a dataset of nearly 5 000 taxa classified according to a predefined set of habitats, the Bugs EcoCodes, which have been derived at by way of their potential usefulness in palaeoenvironmental reconstruction. Thus the habitat descriptions derived from the same species lists using BugStats will not be the same as those of Eyre & Luff (1990), as the Bugs EcoCode classification system allows for the reconstruction of environments beyond those sampled by Eyre & Luff (1990).

This said, palaeoentomological tasks, such as assigning taxa to the Bugs EcoCode classification, would be impossible without such studies of modern ecology. The importance of palaeoecology to modern ecology, on the other hand has long been overlooked by modern ecologists, most of whom have remained remarkably oblivious to the fact that present distributions are the result of millennia of population-environment interactions. The situation does seem to be improving, and a number of modern ecologists are beginning to look at present day species-environment relationships with respect to the fossil record, particularly in terms of climate change (e.g. Davis *et al.,* 2002; Eyre *et al.,* 2006).

Webb & Lott (2006) propose "a habitat-based invertebrate assemblage classification system for assessing conservation interest in England" under the name ISIS. It covers more insect orders and fewer Coleoptera species than BugStats, and is database orientated. The approach to habitat coding appears to be similar, and their description of assigning species to 'assemblage types' seems to be analogous to the assignment of species to Bugs EcoCode habitat groups, Webb & Lott's assemblages not being static lists of species, but rather compounded of species know to occupy a particular habitat[v]. ISIS is designed with modern ecology and conservation in mind, and as such orientated towards the targeted sampling of species in specific environments. These samples are more likely to represent a smaller and more local catchments than the typical fossil sample, which may represent hundreds, if not thousands of years. Bugs EcoCodes and the BugStats system were designed primarily with palaeo-environments in mind, and it would be very interesting to parallel run a number of fossil and modern faunas on both systems.

The treatment of taxa on an individual basis (BugsCEP), or the use of a flexible 'assemblage' definition (ISIS) greatly reduces the risk of problems associated with non-analogue assemblages. With some proxy methods, it can be difficult to reconstruct an environment from a collection of species which are not found together at the present day, and this has lead to a considerable amount of discussion with respect to quantitative reconstructions and so called no-analogue assemblages (see e.g. Bergman *et al.,* 2005; Williams *et al.,* 2001). A flexible approach also allows for greater scope in predicting the future geographical distribution of species, where habitat fragmentation may prevent present habitat assemblages from being maintained.

In palaeoecology we can not always safely assume that, due to taphonomic processes, the faunal assemblage that we investigate is fully proportional to the populations that were living in the environment that the sample represents. These processes may also lead to the exclusion of species that

---

[v] There appears to be some variation in the literature as to the use of the term 'assemblage' – from a specific collection of species found in a sample (mainly palaeoecology), to the species commonly found together and associated by habitat (e.g. in conservation ecology).

were present in the depositional 'death assemblage' (see Kenward, 1975, for a discussion). With all these possible sources of unquantifiable errors, semi-quantitative, broad habitat definitions are often the most appropriate for environments recreated from fossil datasets. Indeed, habitats defined too specifically, perhaps with too much value placed on their quantitative origins, could very well give a false sense of accuracy to palaeoenvironmental reconstructions.

4.2.3.1 Habitat classification and palaeoecology

In palaeoecology and archaeoentomology the number of habitat classification systems are not as numerous as those in modern ecology and entomology. Kenward (Kenward, 1978; Kenward, 2001) and Robinson (e.g. Robinson, 2001) have used both coding and statistics as an aid to interpreting and displaying palaeoentomological data. A number of other authors consistently present environmental reconstructions in terms of well defined habitat types such as 'aquatic', 'woodland' and 'dung' etc. These habitat types are often chosen with respect to environments that the species found are able to indicate, and are based on the researcher's knowledge of modern entomology and ecology.

Any system using fossil data should be firmly based in modern ecology. Koch (1989-92) has classified the majority of the Central European beetle fauna according to habitat requirements on the basis of considerable research. This system is incorporated into BugsCEP, and forms part of the basis for the alternative in-house system (Bugs EcoCodes) used in BugsCEP and BugStats, which is described below (section 4.3). The Koch ecology code system enables extremely detailed descriptions of several aspects of species habitat requirements, and includes a total of no less than 125 different codes. Although this wealth of descriptors has its advantages in terms of giving a more detailed description, it is somewhat difficult to summarise – especially graphically. In addition, its combination of German and Greek derived terms proved extremely difficult to translate directly into English, and the final implementation in BugsCEP is described below (section 4.5). The Bug EcoCode habitat classification system, with only 22 codes, produces diagrams of similar length to a summary pollen diagram, and allows for a more rapid overview of the environmental changes or differences between samples.

Kenward (Kenward, 1978; Kenward, 2001) calculates diversity indices for each of his habitat classes, which although an interesting idea is so strongly dependent on the classification system used, that it may say more about the author's views on classification, or particular interests, than variations in diversity. Standard errors calculated on the data subsets are equally susceptible to classification bias, and as such not particularly meaningful. Other authors have advocated the estimation of standard errors using resampling statistics, which reduce the dependency on population structure assumptions, as a viable alternative (e.g. Hammer & Harper, 2006). A jackknife variant is used later in this thesis to assess the reliability of MCR reconstructions (Chapter 5).

Kenward also derives a wide array of summary statistics for samples, including estimates of standard errors and Fisher's α, from his classified groups. The latter assumes a log-series ranked abundance distribution within the examined communities (Southwood, 1978), and while this may be shown to be true for many fossil assemblages, it difficult to say whether one can then deduce that the inferred living population had the same distribution. In fact, the effects of varying taphonomic processes between samples and sites makes this assumption somewhat debateable. In addition, the fact that it is an untestable hypothesis makes the use of Fisher's α for testing fossil diversity risky. Other diversity measures may be more appropriate, and whether individually valid or not, diversity indicators can provide a useful way of comparing faunas. BugStats provides the facility to compare samples by the extent to which they have species in common, but currently only supports the Bray and Curtis modified Sørensen's coefficient of similarity (Southwood, 1978), as described below (section 4.4.3). The usefulness of this coefficient is demonstrated in the Grande Pile site (Ponel, 1995) and other examples in Chapter 6.

4.2.3.2   Further notes on habitat classification methods

The classification systems described above can be broadly divided into two categories: quantitative and semi-quantitative. The former are based on the summary of numerical data, and often include the ability to estimate errors for derived results, whereas the latter may use a combination of numerical and descriptive data, and have a limited ability to provide indications of the reliability of any conclusions derived from them. Due to the patchy nature of fossil insect data, and the fact that taphonomic problems severely restrict the validity of the required assumptions of certain population structures, and their representation in the fossil assemblages, semi-quantitative methods are more commonly used in palaeoentomology. In fact, there is a considerable danger of over or misinterpretation by the misapplication of quantitative statistical methods to datasets which do not satisfy the basic assumptions of the methods applied. They can, however, be successfully used to interrogate fossil faunas, and ordination and cluster analysis in particular have been used to look at species groups with respect to climate change (Cong & Ashworth, 1997), species associations (Kenward & Carrot, 2006) and flow regimes (Greenwood *et al.,* 2006), among other things.

Finally, although BugStats provides numerically derived reconstruction aids, the assignment of species to the EcoCode reference dataset cannot be considered quantitative, due to variation within the sources used. As a consequence of this, and the fact that the habitats reconstructed have yet to be subjected to thorough independent testing, the BugStats system should be considered a semi-quantitative environmental reconstruction tool.

4.3   The Bugs EcoCode Classification System, and the BugStats Environmental Reconstruction Software

The classification of species has been selective, and species where the records are confusing or unclear, or are described as casual finds in the habitat, were omitted. Classification has been performed using a combination of Koch (1989-92) and the references available in BugsCEP for each species. In general, more than one reference has been used to classify a species, unless the reference is a respected authority on the group, and more than two if references are regionally specific. Reliance on the latter type of information source has been avoided in an attempt to provide a classification system which is valid for most of Europe. This has, of course, lead to a number of generalizations, but hopefully fewer mistakes. The central to north European focus of BugsCEP means that habitats more common in or exclusive to southern Europe, such as the garrigue soft-leaved scrubland, will be under represented. It is also likely that the more continental, eastern European, environments are under-represented.

4.3.1 Bugs EcoCode classification system description

The BugsCEP database contains the facility to assign taxa to any number of seventeen general habitat types or classes (Table 4.1). In addition, five indicator classes are provided for species that are sufficiently stenotopic to be able to imply a very specific habitat type. A few of the latter species may be found in other, similar environments (e.g. dung and decaying vegetation), but are found significantly more often in the primary environment. About 5 000 taxa have been classified at the time of writing, and their distribution amongst the categories is described below (section 4.3.2).

The primary objective in the development of Bugs EcoCodes was to devise a classification system and accompanying software to enable researchers to produce easily visual descriptions of the environments represented by fossil or modern insect faunas. This increases the efficiency of interpretation, providing both a data analysis tool and presentation system in one package. It was also intended that it should facilitate inter-site comparisons, allow researchers, including those from other fields, to easily compare reconstructions, and be a valuable teaching aid.

Table 4.1. List of Bugs EcoCodes, with diagram label and short description. Indicator classes are in *italics*, and * marked classes represent narrow habitats. See text for further explanation.

EcoCode	Label	Detailed definition
BEco1	Aquatics	Living in/on water, in any form. From temporary pools to lakes and rivers.
BEco1a	*Indicators: Standing water*	*Primary habitat in/on pools, ponds, slow flowing water – including temporary ponds, but not species specifically in vegetation and mud or banks of ponds.*
BEco1b	*Indicators: Running water*	*Rivers and/or streams. Species predominantly found in these.*
BEco2	Pasture/Dung	Grazed land of varying form. Includes most dung beetles, including those that are not stenotopic to dung. Mostly open landscape, but may include pasture-woodland when in combination with BEco4.
BEco3	Meadowland	Natural grassland or near equivalents. Open landscape.
BEco4	Wood and trees	Species tied to either the actual wood, trees or the forest/woodland environment. Generally shade tolerant.
BEco4a	*Indicators: Deciduous*	*Specifically deciduous wood or woodland, species not found on coniferous wood except on rare occasions.*
BEco4b	*Indicators: Coniferous*	*Specifically coniferous wood or woodland, species not found on deciduous wood except on rare occasions.*
BEco5a	Wetlands/marshes	Water tolerant but not living specifically in the water. May include mud and bank species, as well as those moss & reed dwellers that prefer permanently wet environments.
BEco5b	Open wet habitats	Hydrophilous shade intolerant species, shingle, beaches etc. and other exposed wet environments.
BEco6a	Disturbed/arable	Any disturbed ground surface, be it by animal, geological or human action. Includes ploughed fields, edges of watering holes, farm yards, glacial margins etc.
BEco6b	Sandy/dry disturbed/arable	Similar to the above, but more xerophilous species. Typifies beach, dune and aeolian landscapes, or ploughed fields on more sandy soils. A more dominant environment in southern Europe than BEco6a.
BEco7a	Dung/foul habitats	A wide category for species that live in decaying, muddy and fetid environments, including compost, wet hay, dung and muddy edges of water.
BEco7b	Carrion*	Animal carcasses of all forms, dry or wet.
BEco7c	*Indicators: Dung*	*Primary habitat dung, or dung essential for reproduction. Includes parasites of other species that live in dung. Majority of species not found in other environments represented by the broader class BEco7a, but some may be found on occasions outside of dung.*
BEco8	Mould beetles of all types	Large part of the typical indoor synanthropic fauna in northern Europe.
BEco9a	General synanthropic	In association with humans, either when outside of their 'natural' geographical range, or in all known records. This term may be geographically specific, and is used in a north European context here.
BEco9b	Stored grain pest*	Pests of stored products.
BEco10	Dry dead wood*	Wood in constructions, but also similar natural environments such as large fallen trees, especially in warmer climates.
BEco12	Heathland & moorland	Heathland and moorland, but may also indicate the under-story of a Boreal forest (see Finnish example in Chapter 6).
BEco13	Halotolerant	Salt tolerant, often coastal or salt marsh tied, but not just NaCl – can be species found on mineral rich ploughed soils or where mineral precipitation is prominent.
Ecto	Ectoparasites*	External parasites of humans and animals.

The habitat classes were defined using a combination of:

1. the usefulness of the habitat class definition in palaeoecology and archaeology
2. the ability of beetle faunas to show the habitat
3. the apparent frequency of use of certain habitat classes by other authors

The classification system was devised by Philip Buckland and Paul Buckland in consultation with other experts[vi], and is based on the data available in BugsCEP, whilst being influenced by existing

---

[vi] In particular, Gunnar Gustavsson helped with the dung beetles, Geoffrey Lemdahl provided a number of comments, and Fredrik Olsson performed extensive testing.

published classifications (Kenward, Koch, Ponel, Robinson, etc., specific references below). Although many authors have classified their faunas by habitat, aside from Koch (1989-92) none appear to have published a definitive list of the species assigned to the classes. Kenward (1978) points out that his species assignments will change with experience and as research on the particular taxa progresses, and it is expected that the BugsCEP classifications will also need to be updated with time. It is therefore essential that the version of BugsCEP/BugStats used is always cited when presenting results. Robinson (2001) presents a useful classification system with 14 species groups, using which he graphically presents interpretations of West Heath Spa (Hampstead Heath) and The Hamel (Oxford)[vii]. Robinson's diagrams show species groups expressed as a percentage of terrestrial Coleoptera, and states that "[n]ot all the terrestrial Coleoptera have been classified into groups." (Robinson, 2001: p129-130). Unlike Kenward, however, he does not present a list of taxa or give any indication as to how (or indeed which) species were classified. A number of the classes have been adapted from Robinson's suggestions[viii], and it is hoped that the use of similar codes will allow a degree of inter-comparability between the systems. A number of the habitat groups used by Ponel (1995) are also similar to those used in BugsCEP, and a comparison with those used in his analysis of the Grande Pile site is provided in Chapter 6.

Many species have habitat ranges wider than the individual Bugs EcoCode classes, and any single species can be found in several classes, that is to say a taxon may represent more than one habitat (see section4.3.2). Indicator classes are slightly different from the general classes, in that a species can only occur in one indicator class, these reflecting the mutually exclusive aspect of stenotopy. A number of general classes, marked with an asterisk (*) in Table 4.1, describe significantly more specific habitats than the others, and could be regarded as indicator classes to a degree. It was, however, decided that marked indicator classes should be only those that are subsets of wider environments represented by the system (e.g. *Standing Water* is a subset of *Aquatics*), although they are not currently treated any differently in the calculations.

To avoid loss of information where taxa have been only identified to generic or higher level (spp., sp., indet.), these taxa are assigned all the codes of the species in the genus, with the exception of the indicator classes. Although this does reduce the 'accuracy' of the reconstruction by weakening the power of the percentage, and possibly presenting a more varied picture of habitats, the authors feel that the information gained is of such value in some situations that this is a valid concept. This feature is of particular benefit when interpreting small or poorly preserved faunas. User discretion is advisable however, for the range of habitats occupied by a genus varies considerably, as illustrated in Table 4.2 (and Figure 4.1). Although a somewhat obvious example, Table 4.2 serves to illustrates the point that knowledge of the organisms involved is essential when interpreting the software outputs. Users may choose to exclude these taxa from calculations using the 'Species level id's only' option described below (section 4.3.3.4).

Table 4.2. Comparison of the habitat codes ascribed to a eurytopic and a stenotopic genus. Note that the apparent degree of stenotopy is a direct function of the code system as well as the ecology of the species.

Genus	Bugs EcoCodes
Otiorhynchus sp. (29 spp.)	BEco4; BEco3; BEco5a; BEco6a; BEco6b; BEco12; BEco9a; BEco13
Agabus sp. (23 spp.)	BEco1 (Aquatics)

[vii] Both sites are in BugsCEP, although The Hamel has only presence/absence data.

[viii] Specifically: *Aquatics*, *Pasture/Dung*, *Meadowland*, *Wood and trees*, *Disturbed/arable* , *Sandy/dry disturbed/arable*, *Dung/foul habitats*, *General synanthropic*, *Stored grain pest*, and *Heathland & moorland*. In total ten of the 22 groups, although it should be noted that similar groups are used by a wide variety of authors.

4.3.2 Bugs EcoCode designations and their implications

Consideration of population biology issues is essential when studying the expression of environmental change as reflected in changing insect faunas. The Coleoptera are an extremely diverse group, and a wide range of population structures are represented, from low spatial density solitary predators (e.g. *Carabus violaceus* L.), to large numbers of highly localised dung feeders (e.g. *Aphodius lapponum* Gyll.), and highly abundant yet host specific leaf feeders (e.g. *Phratora vulgatissima* (L.)). The relative abundance of species in different habitats is of particular importance when reconstructing habitats, and any system that attempts to quantify taxa by habitat will be subject to the implications of these variations. BugStats includes no tools for compensating for differences in species population densities, and this must be considered when interpreting results.

Equally important is the distribution of taxa within the database amongst the classifications available. As can be seen from Figure 4.2, the *Wood and Trees* category is represented by almost twice as many taxa as any other group, and there is considerable variation in the number of taxa per habitat class. This pattern may well be a reflection of the database's central to north European orientation, and is considered to be more of a reflection of reality than an artefact of the classification system, although the latter is of course important. The number of species assigned to indicator classes is naturally low, as species stenotopic to the very specific habitats used in the Bugs EcoCode classification are relatively rare[ix]. The same can explain the low numbers in the ecologically narrower general classes (asterisked in Table 4.1).

Were the insect species listed in BugsCEP randomly assigned to environments in the natural world, the variation in habitat representativeness would be of statistical concern. In reality, insects actively engage in habitat choice by way of their preferences for food sources, climate and vegetation, etc., and some environments support more species than others, be it due to geographical, resource or competition factors (Colinvaux, 1973). An ideal habitat classification system would perhaps reflect these natural inequalities in nature, but they are difficult to quantify, especially for large numbers of species. When interpreting fossil faunas, one must always be aware that not every habitat or area will be represented equally, and that support from other proxy sources is always beneficial, if not essential. It is possible that, with care, the coded ecology data could be used to infer details on the past relative abundance of habitat types in Europe, and it would be interesting to explore the explanation behind the patterns observed in Figure 4.2. Could it be, for example, a reflection of the dominance of woodland environments earlier in the Holocene or is it inherited from a landscape of continuous forest in the mid-Tertiary? Is it perhaps a reflection of patterns of habitat adaptability and range within the Coleoptera of Europe? Unfortunately these questions are outside the scope of this thesis, and must be reserved as the subject of future research.

As described above, taxa may belong to more than one habitat group, although the greater part of BugsCEP taxa are habitat specific enough, or the categories broad enough, for the majority of taxa to belong to only one or two groups (see Figure 4.1). The sample-by-sample breakdown report (see section 3.4.3.3) for a site can be used to see exactly which taxa represent which habitats in a sample, and may be useful in understanding the spread of environments represented. It can also be useful in deciding whether to omit genus level identifications, or possibly low abundance taxa from the calculations.

[ix] However, of the c. 3 000 species in BugsCEP with Koch (1989-92) classifications, approx. 50 % of the species are considered stenotopic in his considerably more detailed system. This highlights the influence of the classification system on observed patterns, and that the interpretation of classified data is not as straight forward as many might think.
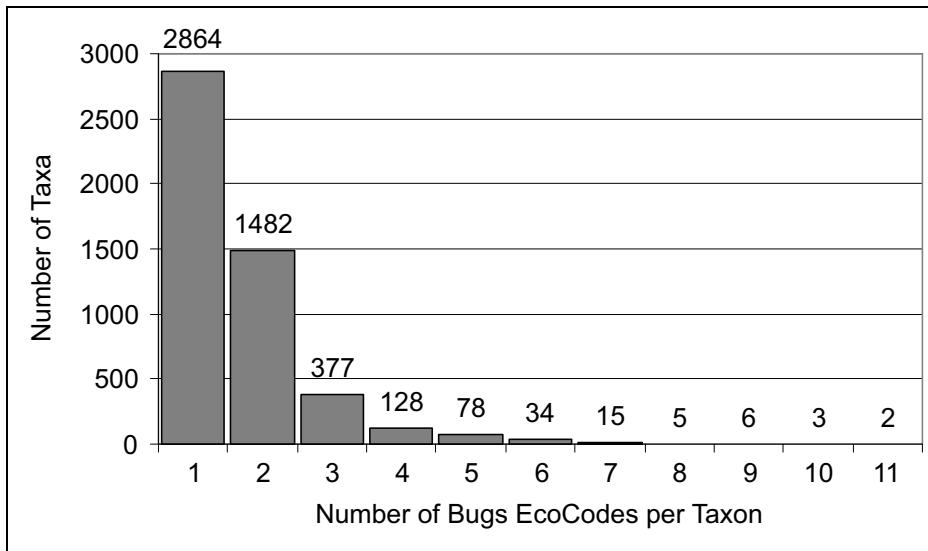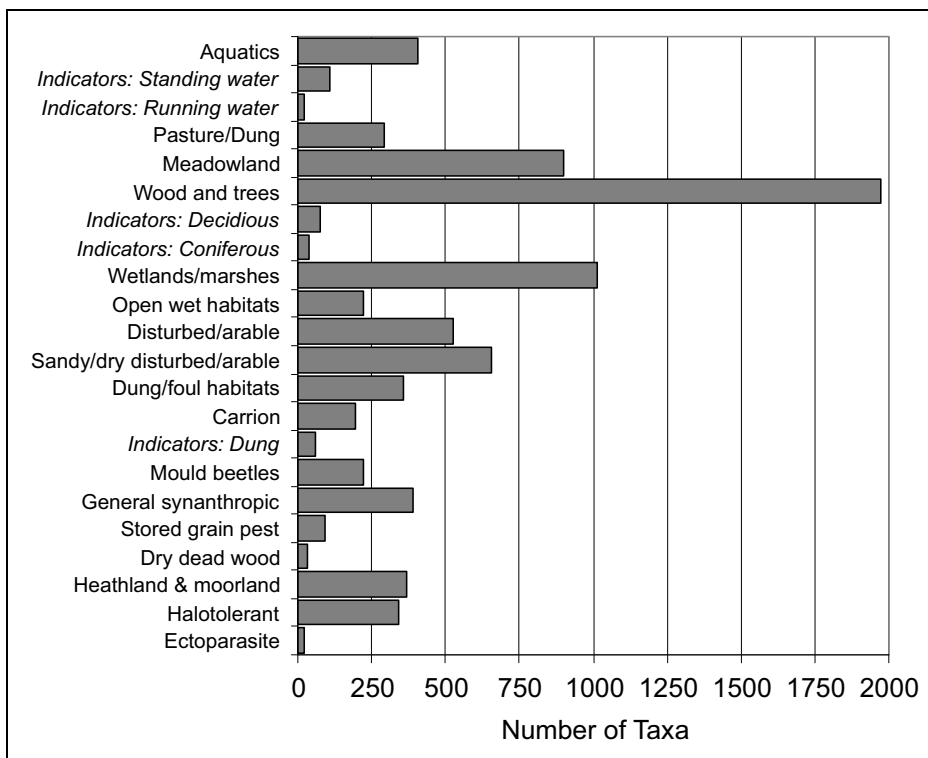
Figure 4.1. Number of Bugs EcoCodes per taxon.

Figure 4.2. Number of taxa per Bugs EcoCode habitat class. Taxa can be present in more than one general class, but only one 'Indicator' class.

4.3.3 EcoFig calculations, transformation and standardization

The Bugs EcoCode classifications of species found at a site are used as the source data for the construction of EcoFig diagrams, which are the primary graphical output of BugStats. On initiation of EcoFig calculation for any countsheet, BugStats tallies the EcoCodes for each species in each sample, abundance weighted if required, and produces a raw data table which it exports to MS Excel. A sum of these tallies (SumRep = Sum of environmental representations, or the total of all counts for a sample), sample abundances (abund) and number of taxa (nspec) counts, along with sample names and other details are appended to this. The results are then recalculated according to one of the standardization

models given below, and exported as a new worksheet in the same MS Excel file, after which the results are graphed as a series of bar charts (Figure 4.3, explained in section 4.3.3.3). BugStats uses a programmed intelligent scaling system to avoid problems with MS Excel's automatic scaling routine, and to ensure that all non-sum graphs have the same scale. Figure 4.4 summarizes the above process as a flow diagram.

4.3.3.1 Excluding taxa not identified to species level

The user may choose to omit taxa that have been only identified to a higher taxonomic level (sp., spp., indet.) from the calculations. Doing so may produce a diagram that more accurately reflects the immediate environment of the sample, although the true implications may be dominated by taphonomic issues.

4.3.3.2 Logarithmic transformation

Transformation by the function $\ln(n+1)$ [$=\log_e(n+1)$] is available by selecting the appropriate check box on the BugStats screen. The natural logarithm (as opposed to logarithm to base 10) transformation is provided as recommended by Jongman *et al.* (1995) for compensating for a logarithmic response of species abundances to environmental variables. It is up to the discretion of the user as to whether to apply this, and it can be applied to non-abundance weighted calculations as well as abundance weighted. The option is mainly provided for comparability with other work, although it may not be entirely relevant for a system based on classified data. Taxa are irregularly distributed to the classes, there are a varying number of classes per taxa, and it is debatable as to whether uniformly applying a single transformation over taxa that may have differing population structures is appropriate. The degree to which the habitat enumeration process enhances or diminishes the population structure of a sample is highly dependent on the degree of stenotopy of the fauna, as well as the natural population densities. One can safely assume that different samples may contain species at different distances from their distributional optima, and thus naturally occurring at different frequencies in the samples and providing a weaker signal of their preferred habitats. To log transform this data, and thus decrease the prominence of highly abundant taxa and increase the importance of rare taxa, could lead to the dilution of patterns important for explaining the fauna's environmental implications. In addition, the taphonomic issues that characterize palaeoentomology may cause unpredictable artefacts in the nature of the effect of the transformation on the data.

This said, the addition of more complex models based on population biology into the software, even if only for the sake of experimentation, is not precluded. Where the population structure of the taxa involved is well understood, for example, it may be desirable to differentially weight the species abundances, perhaps in a way similar to that used in weighting pollen production in palynology (e.g. Prentice, 1985). Also, in certain reconstruction environments it may be desirable to apply known models of population structure, for example, to transform the count values in a way which compensates for naturally superabundant and low abundance species within that particular environment. There will always be the risk, however, of modelling and transformation adding complex, if not unpredictable bias to the system. More experimental work on the representativeness of fossil insect faunas is required before we can begin to apply such corrections. Lessons are undoubtedly to be learned from the work on pollen representativeness (e.g. Hicks, 1993) and landscape reconstruction (e.g. Broström, 2002; Sugita *et al.*, 1999), even if the organisms are significantly different.

Note that transformations are applied after summing the EcoCodes and before standardizing (see Figure 4.4), and thus do not transform the actual abundance data, but rather the habitat group sample sums, or the abundance weighted sums.
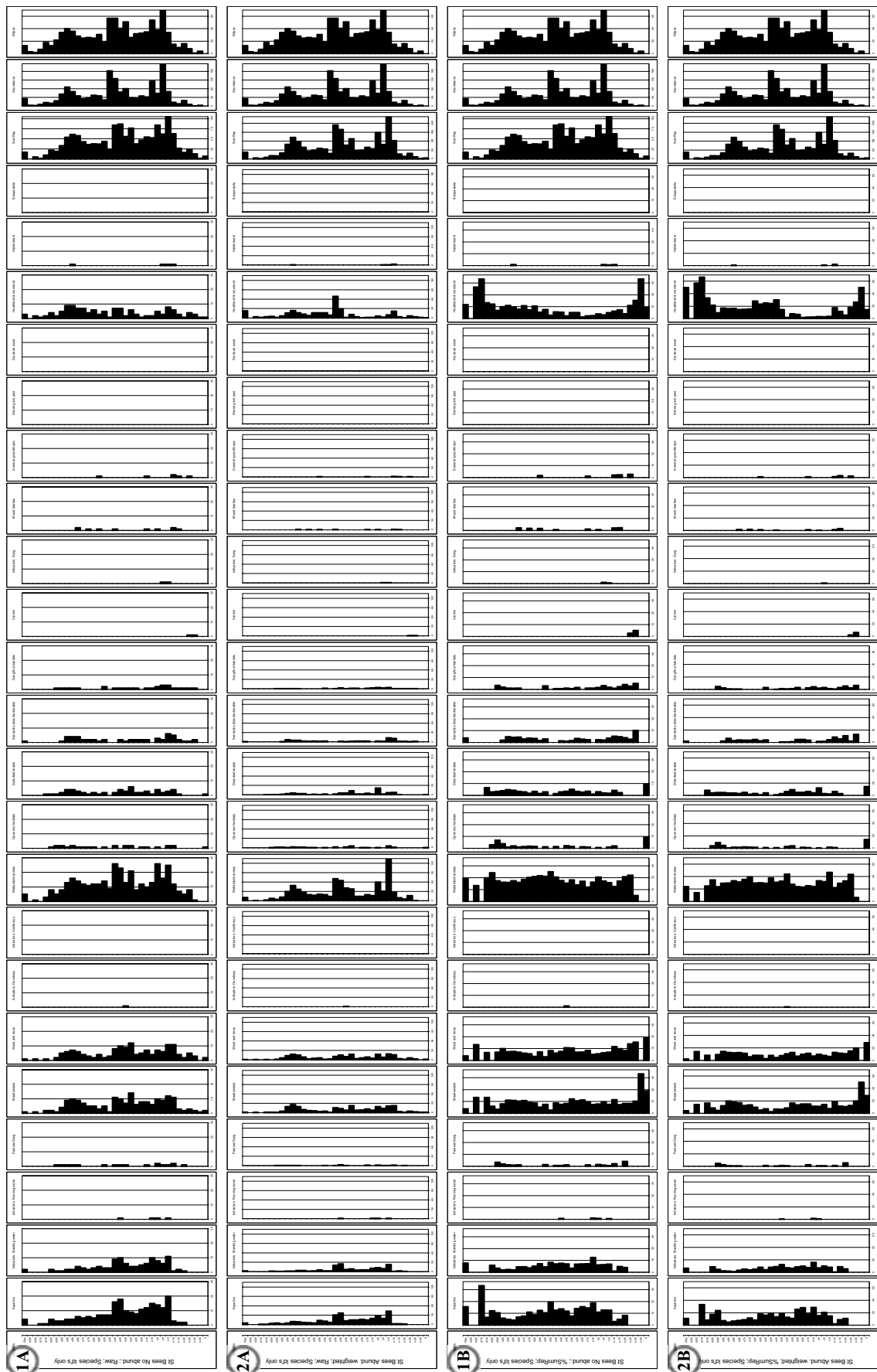
Figure 4.3. *Unmodified* BugStats EcoFigs for Saint Bees (Coope & Joachim, 1980), note that the basal sample is at the top of each diagram. Settings, from the left: 1A: No abundance, Raw; 2A: Abundance weighted, Raw; 1B: No abundance, %SumRep; 2B: Abundance weighted, %SumRep (see 4.3.3.3).
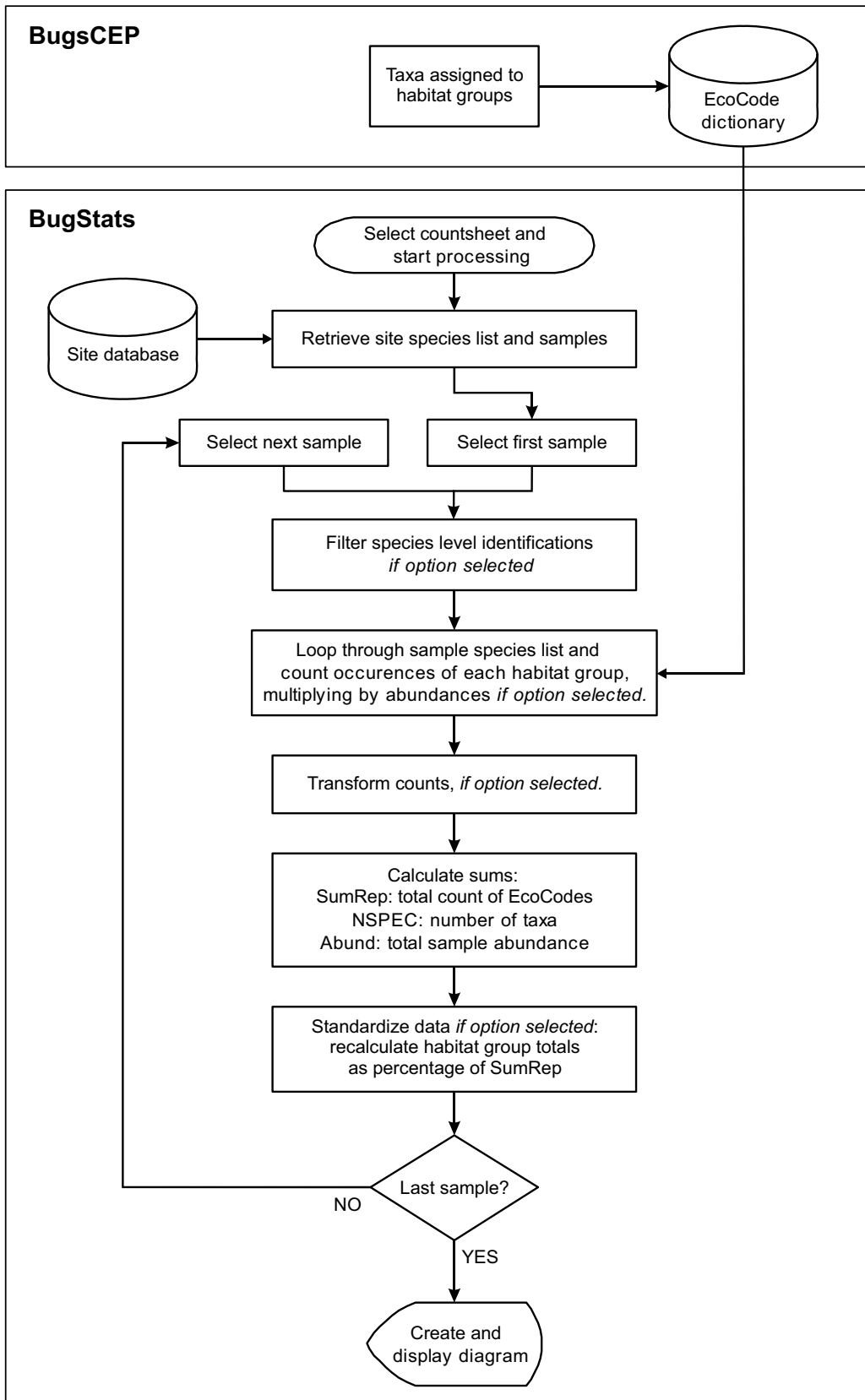
BugsCEP

```
┌─────────────────┐
│ Taxa assigned to │ ──────▶ ⬭ EcoCode
│ habitat groups   │            dictionary
└─────────────────┘
```

BugStats

⬭ Select countsheet and
start processing

⬭ Site database ──▶ Retrieve site species list and samples

Select next sample   Select first sample

Filter species level identifications
if option selected

Loop through sample species list and
count occurences of each habitat group,
multiplying by abundances *if option selected.*

Transform counts, *if option selected.*

Calculate sums:
SumRep: total count of EcoCodes
NSPEC: number of taxa
Abund: total sample abundance

Standardize data *if option selected*:
recalculate habitat group totals
as percentage of SumRep

Last sample?
NO
YES

Create and
display diagram

Figure 4.4. Flow diagram illustrating the sequence of events that create an EcoFig diagram from site data.

4.3.3.3 Standardization

The often large variation in the number of species and individuals between samples can make it difficult to compare the raw habitat group counts from different samples. A form of standardization needs to be applied to make the habitat group sums comparable, such as is commonly seen in percentage pollen diagrams (see Berglund & Ralska-Jasiewiczowa, 1986). The simplest form of standardization is perhaps a percentage or proportion, but due to the fact that taxa may be assigned to more than one habitat in the Bugs EcoCode system, and that either numbers of taxa or individuals may be used in the calculations, several possibilities are available. Both standardized and raw outputs are available from BugStats, and both should be examined so that no patterns are missed.

After experimenting with various forms of recalculation, at varying degrees of complexity, it was decided that the most useful options to provide would be the following:

1 – No abundance (taxa only)

2 – Abundance weighted

A – Raw class count data

B – Counts expressed as a percentage of sample sum of counts (%SumRep)

These options can be combined, resulting in the four alternatives below[x]. Examples are provided to illustrate the effects on sample calculation results, which are given in **bold**, and full EcoFigs are shown for each alternative in Figure 4.3.

1A: No abundance; Raw

Non-abundance weighted values are simply a count of the number of taxa in each habitat class. Since taxa can occupy more than one habitat class, the sum of counts (SumRep) will generally be greater than the number of taxa found in the sample (NSpec).

Example: If habitat group *Wood and Trees* (BEco4) is represented by 4 taxa in sample S5, then this class receives the count **4** for sample S5.

2A: Abundance weighted; Raw

Abundance weighting multiplies the count of each taxon in a habitat group by its abundance in that sample. Note that since taxa can represent more than one environment the abundance weighted sum (abSumRep[xi]) will normally be greater than the sample abundance (Abund).

Example: If habitat group *Wood and Trees* (BEco4) is represented by 4 taxa in sample S5, with the abundances 4;6;13;1, then this class receives the abundance weighed count of **24** for sample S5.

1B: No abundance; %SumRep (standardized)

Non-abundance weighted standardized values are calculated, for sample S and class C, as the raw counts divided by the sum of all habitats counts for sample S, multiplied by 100 (see Equation 4.1). This is referred to as %SumRep, or the percentage of the sum of environments represented.

---

[x] Note that this is not the order that the options are shown in BugStats, but rather an order that simplifies explanation.

[xi] In the software this is simply referred to as SumRep, and abSumRep is only used in the text here for clarity.

Equation 4.1. Calculation of standardized, taxa only, EcoFig class values.

$$\%\text{SumRep} = \frac{RepCount(S,C)}{SumRep(S)} \times 100 = \frac{RepCount(S,C)}{\sum_{C=1\,to\,n} RepCount(S,C)} \times 100$$

where:

$RepCount(S,C)$ = the number of taxa representing habitat class C in sample S.

$SumRep(S)$ = the total number of taxa-environment representations in all classes for sample S, or the sum of *RepCount* for all classes for sample C.

n = the number of habitat classes.

Example: If habitat group *Wood and Trees* (BEco4) is represented by 4 taxa in sample S5, and the total sum of taxa counts for all environments represented by taxa in sample S5 is 30, then this class receives the standardized value (4/30x100=) **13.3** for sample S5.

2B: Abundance weighted; %SumRep (standardized)

Abundance weighted standardized values are calculated, for sample S and class C, as the abundance weighted habitat group counts divided by the sum of abundance weighted habitat group counts for sample S, multiplied by 100 (see Equation 4.2). This is referred to as abundance weighted %SumRep.

Equation 4.2. Calculation of standardized, abundance weighted, EcoFig class values.

$$\%\text{SumRep} = \frac{Abundance(S,C)}{abSumRep(S)} \times 100 = \frac{Abundance(S,C)}{\sum_{C=1\,to\,n} Abundance(S,C)} \times 100$$

where:

$Abundance(S,C)$ = the sum of abundances (= total number of individuals) of all taxa in sample S that represent habitat class C.

$abSumRep(S)$ = the sum of abundance weighted environmental representation counts over all habitat classes for sample S.

n = the number of habitat classes.

Example: If habitat group *Wood and Trees* (BEco4) is represented by 4 taxa in sample S5, with the abundances 4;6;13;1, and the total sum of abundance weighted counts for all environments represented in sample S5 is 80, then this class receives the standardized value (24/80x100=) **30** for sample S5.

4.3.3.4 *EcoFig diagram creation*

BugStats provides output in numerical and graphical form. The details of the options available for customizing EcoFigs are given in section 3.4.3.2, and will not be repeated here. EcoFigs are constructed from a series of horizontal bar charts that represent the habitat classes, along with a figure showing sample names and additional charts for the sample sums. Each habitat class chart is given the abbreviated title of the class that it represents, and the sum columns are labelled accordingly. The first

chart, with sample names, also includes a vertical text banner including any title specified by the user, appended with details of any standardization applied.

The unmodified output from BugStats for the site Saint Bees (Coope & Joachim, 1980) is shown in Figure 4.3. Each diagram was created by each of the combination of options described above (section 4.3.3.3), and with the 'Species level id's only' option selected. The diagrams are not ready for publication in this state, but are presented here to illustrate the real output of BugStats[xii]. Using the instructions provided in the help files, any diagram can be copied and pasted into a vector based graphics package (e.g. CorelDraw, Adobe Illustrator) and edited into a more publishable form by increasing text sizes, removing empty charts and truncating charts that do not use the entire scale. Notice that the last two charts on each diagram row are identical, and show the number of individuals and species in each sample. Samples are arranged in the order that they are stored in the countsheet, in this case with the basal sample at the top, which may not be ideal in all circumstances. An option to sort samples in diagrams will be included at a later date, and for now users should resort the samples on the 'PctResults' worksheet in the results file, before copying the diagram into a graphics package.

The bars in each habitat class chart represent the relative or absolute prominence of that habitat type in the sample. One can clearly see that in the non-standardized diagrams (Figure 4.3, 1A and 2A) the bars in the majority of groups follow the relative changes in number of species or abundance. Standardization (Figure 4.3, 1B and 2B) essentially rescales the habitat bars for each to be proportional to the total number of habitat indications in the sample, and thus allows samples to be compared more reliably. In particular, one can see that the gradual trend in the left most bar chart, *Aquatics*, is almost lost once standardized. The standardization suggests a much greater availability of aquatic habitats in the upper parts of the chart than the raw values do. In the majority of the more strongly represented habitats the peaks and troughs are evened out by standardization, with the exception of *Heathland & moorland*, which gains significant peaks at the top and bottom of the sequence. Note that the habitat chart scales are different on each diagram row, due to the auto-scaling routine.

Chapter 6 provides more detailed explanations of the use of EcoFigs and the implications of standardization, including comparisons with previous research and diagrams modified to publication standards.

4.3.3.5   Sample by sample EcoCode report

This report simply returns the Bugs EcoCode data, sample by sample, for the taxa at a site, along with the total number of each taxa in each sample. It may be useful to consult this report when interpreting the EcoFigs, or deciding which calculation options to use. An example is shown in section 3.4.3.3.

4.3.3.6   Alternative standardization possibilities

Other standardization techniques could easily be implemented if they were found to be useful. Rarefaction, for example, could be used to estimate species numbers based on the relative abundance of species in the samples, and has been used in a number of studies using pollen (e.g. Birks & Line, 1992; Odgaard, 1999), diatoms (e.g. Wolfe, 2003) and chironomids (e.g. Nyman *et al.*, 2005). The method appears to have been extremely rarely used in palaeoentomology, if at all[xiii], but is more common in modern entomological studies (Krebs, 1999; e.g. Lassau *et al.*, 2005). As the usefulness of the method is dependent on the size of the smallest sample, it will be problematic in palaeoentomology as most sites include at least a few very poor samples.

[xii] Differences in the Windows desktop resolution settings (DPI-settings) mean that font size instructions are not always predictable. A smaller font size has therefore been used on diagrams to reduce the risk of the unpredictable rescaling of diagrams which can occur in MS Excel if titles do not fit on one line. The same applies to the x-axis scales and sample names.

[xiii] A literature search failed to retrieve any example.

A deceptively obvious, and previously often used, candidate for standardization is by original sample weight or volume. There are several important issues to consider, however, including accumulation rates and factors that could naturally affect the relative abundance of species. A constant accumulation rate implies that, for a hypothetically constant environment, the rate of deposition of insect fragments should be the same throughout the profile. In reality, as sedimentation rates vary there is unlikely to have been a constant rate of deposition of insect fragments, and this affects the validity of weight/sample size standardization. It is also well established that both climate and environment type and complexity effect biodiversity and the abundance of species (Lassau *et al.*, 2005; Colinvaux, 1973), and the richness of samples should reflect this. For example, areas with either extremely warm or cold climates are generally less species rich than temperate climates, and would be expected to produce fewer fossils per unit volume of original sample.

The alternative of using the processed volume or weight of the sample is not viable, as this is notoriously difficult to measure without drying the sample and causing damage to the specimens. Its usefulness is also debatable due to the above concerns, and the large variation in the effectiveness of floatation and the materials that are retrieved, dependent upon sediment type.

4.3.3.7   *Final note on the use of BugStats options*

The influence of numbers of individuals on the results of BugStats calculations is neither simple nor linear, as can been seen by plotting abundance weighted minus non-weighted results for the same site. Some habitat classes are clearly more sensitive to abundance than others (see examples in Chapter 6), and this is undoubtedly both a reflection of the coding system and the population structure of the species involved. Careful inspection of both the raw and standardized results, especially with abundance weighting, in combination with more comprehensive sedimentological analyses, could help investigations into the reasons behind the varying frequency of fossils in different sedimentation environments. Examination of the standardized data can provide a good picture of relative environmental changes, independent of differences in the numbers of species or individuals in the samples. The raw, non-standardized results, along with the sample-by-sample breakdown report, can then be used to assess the importance of the actual differences in abundances.

4.3.4 Known issues with EcoCodes

Aside from the possibility of mistakes in the assignment of taxa to EcoCodes, there are a few particular issues of concern which should be considered when using the EcoFig functions in BugStats.

4.3.4.1   *Geographical variation in habitat preference*

There is evidence that insect species may exhibit variation in habitat specificity throughout their distribution. This may purely be a reflection of increased rarity, and therefore insect captures and accumulated habitat data, towards the limits of a species' range. A number of xylophagous species, ofr example, appear to show increased restriction in terms of tree species at their natural limits, and Buckland (1975) provides a discussion of the death watch beetle, *Xestobium rufovillosum* Deg. as an example. Although a few of the factors affecting apparent habitat specificity can be put down to taxonomic issues (e.g. *Agabus bipustulatus* vs. *Agabus solieri*, Drotz, 2003), a large number of species are geographically more eurytopic than local studies would often seem to indicate, as examination of the biology data in BugsCEP reveals. Undoubtedly, some examples are the result of the restricted definition of habitat groups and the wealth of habitat variation in nature. Other examples, particularly concerning species tied to specific host plants, are less debatable. In constructing the EcoCode database we have avoided assigning indicator status to species where there is evidence of variation in habitat specificity, even if they are used as indicators in some regions. This allows us greater claim to a system that is applicable to the entire central and north European region, and possibly the rest of Europe as well, at the risk of disappointing researchers working on regional studies where the species appear to be stenotopic. In some cases, species approaching their distributional limits may appear to be

more habitat specific than in core areas, and this may well be the case for some of them, but it could equally be a reflection of infrequency of collection and collectors. For example, *Dryophthorus corticalis* (Payk.) is known in the UK from finds in a few oaks, often in association with the ant *Lasius brunneus* (Latreille), in Windsor Forest (Donisthorpe 1939; Hyman, 1992). On the continent, however, it is much more catholic in its choice of trees, appearing in conifers in Sweden (Koch 1992). The fossil record of *D. corticalis* suggests a much more widespread distribution in the UK earlier during the Holocene. There are undoubtedly a number of other species where the limited availability of modern records gives an impression of stenotopy that is unreliable. Careful examination of the fossil record, and comparison between international records may help to highlight these species and define key target areas for future research. Many of these species are likely to be rare in some regions, and this work could help in the collation of Red Data Book lists and their application to regional conservation strategies.

4.3.4.2 Indicators and standardization

The indicator species are currently included in the sums used in standardization. A result of this is that high numbers of indicator individuals/species will result in a relative percentage decrease in the other classes over the rest of the sample, potentially more so where abundance weighting is used. This may lead to interpretative problems in that the narrow indicator habitats may seem more prominent than they are in reality, and a future aim is to add the option to calculate these outside of the general percentage. On the other hand, including the indicators in the overall percentage gives a good signal for the relative prominence of the indicator species in relation to the overall assemblage. It is possible that calculating the indicators as percentage of total non-indicators would be an option, but more work on modern faunas is necessary to evaluate the usefulness of such a feature. However, there is always a danger of over-working the calculations, and by adding more possibilities for users to affect small variations in the results the inter-comparability of investigations may be jeopardized.

4.3.4.3 Indicators and diversity

During the classification of species, it became apparent that stenotopic standing water species are far more numerous than those specific to running water (Figure 4.2). The effect of this in BugStats outputs is that indications of standing water are far more likely to be discovered than those of running water. This is considered as a biological reality for the moment[xiv], and is not compensated for in any way, as to do so could be misleading for general use. It is possible that some form of differential balancing/weighting system could be used for studies into water beetle dynamics, but the many assumptions involved in population biology may not be equally applicable to fossil assemblages due to the effects of taphonomic processes. More work on the representativeness of fossil assemblages in aquatic environments, in collaboration with ecologists, is necessary in order to evaluate these issues.

4.3.5 Correlation coefficients

The Bray and Curtis modification of Sørensen's coefficient of similarity (Southwood, 1978; Krebs, 1999) can be used in BugsCEP to compare the faunas of samples at a site (Equation 4.3). The use of correlation coefficients has been discussed elsewhere by the author of this thesis (Buckland, 2000), and they are frequently used for inter-sample comparisons and as the basis for the construction of dendrograms. An enormous variety of equations are available, each with their own relative advantages and disadvantages in different contexts (see Southwood & Henderson, 2000; Krebs, 1999).

[xiv] Although running water species are at greater risk of being transported into, and then collected from, areas of standing water.

Equation 4.3 Modified Sørensen's coefficient of similarity (C_N)

$$C_N = \frac{2jN}{(aN + bN)}$$

where:

    jN = The sum of the lower abundance values where species are common to both samples

    aN = Total number of individuals in sample a

    bN = Total number of individuals in sample b

The modified Sørensen's coefficient of similarity, which is the compliment (1-B) of the Bray-Curtis coefficient of dissimilarity (Krebs, 1999), uses abundance data and is relatively unaffected by rare taxa. The insensitivity is useful in that rare species are quite common in fossil assemblages, and they tend to be more valuable as indicators of particular environments than useful in statistical operations. Although the coefficient is sensitive to super-abundant taxa, the equation does attempt to balance this by summing the lower values where the same taxon is found in both samples.

The correlation coefficients module creates a results matrix in an MS Excel file, where the similarity of each and every sample may be compared (see section 3.4.3.5 for instructions). If the samples are from a stratigraphic sequence, then the matrix can be used to assist in the definition of faunal groups, as illustrated by example in section 6.2.5, and dendrograms may be created using third party software. If the samples are not stratigraphic in nature, then the matrix may be sorted, or cluster analysis employed, to improve its ability to reveal clusters of similar samples. This may be especially useful in archaeological sites with large numbers of samples, such as that examined by Perry *et al.* (1985) at Stóraborg in Iceland.

Significance testing of coefficient calculations will be introduced in a later version of the program, including the use of resampling techniques as discussed in Chapter 5.

4.4   The mechanics of BugStats

This section outlines the flow of events, in programmatic terms, that lie behind the functions of BugStats. It is primarily orientated towards those wishing to develop similar systems themselves, perhaps for other proxies, as a guide to this particular set of solutions. The program implementation of the methods described above is relatively simple, being generally a combination of sample loops and mathematical operations. Although there is room for modular improvement (there is a degree of repetition and redundancy in the code), the code is structured such that further algorithms or manipulations can easily be added, and the existing sample loops and functions can be adapted to new applications. For the sake of simplicity, the abundance data for a countsheet will simply be referred to as a "countsheet", but in reality the data are in linear, normalized form and must be manipulated to form a cross tabulation, as described in section 3.1.3.3. Figure 4.4 should be referred to for clarification of the flow of events from the point of view of the user.

4.4.1 EcoFig calculations and diagram creation

EcoCodes are enumerated for the taxa (or just the species, if required) according to the selected weighting (abundance or taxa). This task is performed by two SQL queries generated on the fly, using the selected options and countsheet. The first query counts the EcoCodes for all the taxa in each sample, and the second cross-tabulates this into a tabular form, calculating the sums of the sample

counts (or sum of environmental representations: SumRep). Sample summary data are collated in an additional query. These tabulations are read into two arrays in VBA and combined. At this point a template MS Excel file is created for exporting the data in the location chosen by the user, and the standard reference and method headers exported. The raw counts are now log transformed if this option is selected (and the sample SumReps recalculated accordingly), and the raw results are exported into the 'RawResults' worksheet in the MS Excel file. If a %SumRep option is chosen then the raw results are standardized, and the results exported into the 'PctResults' worksheet[xv], or the raw results exported again if no standardization is chosen.

The EcoFig diagrams are created from the standardized results, which are still held in memory after export. If the standardization option was not selected, then the raw results are in their place. The data are examined for peaks, and the best scale range and interval values for all EcoCode bar charts calculated. The summary charts scaled individually using the same routine. BugsCEP then takes control of MS Excel and, in a somewhat complicated and repetitive routine creates, formats and positions a horizontal bar chart for each data column in the 'PctResults' worksheet. The first chart, which contains sample names, is created by 'tricking' MS Excel into creating an empty chart with an invisible negative scale[xvi]. This is all done without hiding the MS Excel application, so that users are fully aware that the system is busy[xvii]. The MS Excel file is then closed, and the user informed of completion.

4.4.2 EcoCode reports

Data are compiled for the EcoCode Report by a complex query, extracting ten columns of data from seven related tables, which is built on the fly from the selected countsheet. These data are used as the source for a report which groups the data by three nested levels: site, sample, and taxonomic code, with EcoCodes being displayed per taxonomic code within each sample. The report is presented using the MS Access preview function, with a custom menu allowing printing or export to MS Word or Excel.

4.4.3 Coefficient calculation

The selected countsheet data are collated with a cross-tab query and read into an array in VBA, where they are log transformed if required. Sample names, number of taxa, and abundance sums are collated in another query which is also read into an array. The sample names are used to construct a symmetrical array to hold the results matrix (cf. trellis diagram). The countsheet data are then fed through the coefficient calculation routine, one sample at a time, in which the current sample is compared with all others. The coefficient value resulting from each sample-sample comparison is saved to the appropriate cell in the results matrix array. An MS Excel file is then created and the results exported, and the user is given the option of opening the file on completion.

The calculation routine code is already prepared for the inclusion of further coefficients, including Jaccard's, Sørensen's (unmodified), and Kulezynski's (Southwood, 1978).

4.5   Further developments and additional methods

Due to the object orientated nature of the MS Access and VBA developer environment, the addition of other methods would require relatively little alteration to the user interface – simply the addition of more buttons, check boxes or menus. Check boxes could be used to specify further standardization, transformation or other manipulation options to be applied in the creation of EcoFigs. As mentioned

---

[xv] 'PctResults' = Percent results, for simplicity in naming, and a legacy of previous versions.

[xvi] Such creative formatting is often necessary in order to produce the desired chart in MS Excel, and the MCR graphing routine performs a similar trick.

[xvii] It also makes the software look more impressive!

above, the modular nature of the code make additions relatively easy, and graphing routines could be tailor made to fit the needs of specific methods. MS Excel's graphing engine is potentially a limiting factor, however, even if the default graphs can be customised considerably, and attempts to 'trick' it into producing dendrograms in particular proved too time consuming. An alternative would be to directly employ Microsoft's graphing engine, which is more flexible but would involve a more complex export routine.

Theoretically, one could feed the results of a classification based (semi-)quantification into ordination, and see if any groupings appear. This could, however, increase the chances of produce artefacts caused by the classification system and ordination method, and thus increase the risk of misinterpretation. It would also complicate the analysis even more, drawing away from the goals of simplicity and transparency that were laid out in the previous chapters of this thesis.

It was initially intended that a system be developed to produce diagrams similar to EcoFigs using Koch's (1989-92) ecology classification system. This would allow for more detailed pictures of the specific ecology and habitat of the beetles found in samples to be built, and hopefully more complex questions on the past to be answered. For example, can one see the effects of coppicing through changes in the proportions of leaf feeders and other woodland species as the woodland is managed? The system, with 125 categories in six groups (Table 4.3), is so comprehensive as to make its direct use as a summary system problematic. Some form of interactive diagram system would be ideal, where sublevels and particular details could be viewed on demand. This is not practical for traditional publications, however, and other forms, including the selective output or combination of classes, need to be investigated.

Table 4.3. Koch (1989-92) ecology classifications as implemented in BugsCEP. Translation provided by Paul Buckland, assisted by Eva Panagiotakopulu. Classes in **bold** were added by the translators to improve the usefulness of the system in archaeology.

Group	Categories			
Ecology	agaricolous	deserticolous	nidicolous	silvicolous
	akrodendric	detriticolous	nivicolous	sphagnicolous
	aquatic	floricolous	paludicolous	steppicolous
	arboricolous	fungicolous	parasitic	stercoricolous
	arenicolous	graminicolous	petricolous	succicolous
	arundicolous	herbicolous	phyllicolous	terricolous or subterranean
	arvicolous	humicolous	phyllodetriticolous	
	boleticolous	lignicolous	phytodetriticolous	torrenticolous
	cadavericolous	limnicolous	planticolous	xylodetriticolous
	campicolous	linicolous	polyporicolous	zoodetriticolous
	cavernicolous	microcavernicolous	praticolous	
	corticolous	muscicolous	ripicolous	
Food Condition	holoprobic	mesoprobic	oligoprobic	
Food Dependency	ectophagous	monophagous	polyphagous	
	endophagous	oligophagous	xenophagous	
	merotopic	omnivorous	xenophilous	
Food Nourishment	algophagous	cortivorous	muscophagous	phytophagous
	aphidophagous	entomophagous	mycetophagous	pollenophagous
	blastophagous	fructivorous	myrmecophagous	rhizophagous
	carpophagous	helminthophagous	necrophagous	saprophagous
	caulophagous	kreophilous	oophagous	sporophagous
	cecidophagous	lichenophagous	phloeophagous	xylophagous
	coprophilous	molluscophagous	phyllophagous	zoophagous
Habitat Range	euryhygric	stenohygric	**Typically synanthropic**	
	eurythermal	stenothermic		
	eurytopic	stenotopic	ubiquitous	
	Facultative synanthropic	**Strong synanthropic**		
		synanthropic		
Habitat Type	acidophilous	heliophilous	petrophilous	thermophilous
	amylophilous	hygrophilous	pholeophilous	tixophilous
	apoidephilous	krenophilous	psammophilous	trogophilous
	chromophilous	mycetophilous	rheophilous	tyrphophilous
	coprophilous	myrmecophilous	rodentophilous	xerophilous
	halophilous	necrophilous	saprophilous	xylophilous
	halotolerant	osmophilous	silicophilous	

4.6   Conclusions

Although the Bugs EcoCode and BugStats system is in its early days, its usefulness in aiding palaeoenvironmental reconstruction can be demonstrated (see Chapter 6 and Buckland [*et al.*], 2005), and preliminary testing suggests that the standardization method is effective (see Chapter 6, 20 000 year dataset example). It has clear advantages over the use of ordination where inter-site comparisons are to be made, although it may not immediately provide for the subtle investigation into environmental gradients that the former does, being restricted to the 22 habitat groups of the reference dataset.

The application of BugStats to modern reference studies is essential if it is to be refined. This would allow us to test more thoroughly the ability of the system to resolve habitats, and possibly build in more statistical functions to improve environmental reconstructions. Three types of environment are of particular interest:

1. Specific localities where indicator species are expected to be present (e.g. deciduous woodland, farm yards, lakes, fast flowing streams).
2. Ecotones, the areas of overlap between the locality types above (e.g. woodland edges, lake shores, the moving boundary between the Arctic and the taiga, boundaries created by renaturalisation (re-wilding) projects).
3. Transitional environments with implications for landscape and heritage management, conservation and sustainable development (e.g. natural grazing projects, nature reserves, industrial clean-up operations).

In addition, further studies on fossil sites (including archaeological), perhaps with the possibility of reinterpretation of these, are needed to build up a database of semi-quantitative reconstructions of Quaternary environments. As is often the case in research orientated software design, there has been little time to run the software on the data available. There has not been time to run BugStats on all sites in BugsCEP, but the process could easily be automated. In addition, the application of coefficients of comparison to the BugStats output data, that is to say the comparison of samples by their environmental reconstructions rather than assemblages, could prove an interesting exercise.

The software should probably still be considered a work in progress, and its true reliability over a wide range of fossil and modern sites is yet to be proven. Future work will focus particularly on a comparison of published interpretations and those assisted by BugStats, and the validation of the method through the analysis of modern sites.

The real limits to enhancement, aside from issues of time and funding, may in fact be the evaluation of what one practically *needs* to include in BugsCEP. Should, for example, common, more mathematically advanced methods be incorporated when there are existing software packages that perform them? At the moment, BugStats performs calculations that no other package provides, and can be considered as a specialist tool for palaeoentomology and entomology. The addition of further, standard methods may be an unnecessary adventure into the realm of generic statistics packages. On the other hand, BugStats has the advantage of being connected to the BugsCEP database – the only database of its kind, and with fossil insect data available for immediate analysis, without the need for the export and import needed to get it into another package. These questions are things that can only be answered by BugsCEP users, with whom, and with time and a degree of networking the future path for BugStats will be developed.