
1 Introduction

1.1 Aims of the Thesis

The work behind this thesis is essentially that of methodological development – more specifically the development of software to act as a research and teaching tool for palaeoentomology and ecology. BugsCEP, as the Bugs Coleopteran Ecology Package is abbreviated, has been developed to this end, and is described here along with a number of case studies and worked examples to illustrate its purpose and evaluate its usefulness. The analysis of fossil insect remains is a valuable method in the study of past environments and climates, and important in both environmental archaeology and Quaternary geology research. The software described here is developed in such a way as to also be of use to landscape ecologists, environmental scientists and entomologists. Whilst the database currently centres upon Coleoptera (beetles), it also provides a framework for expansion into other insect groups of use within palaeoecology and environmental archaeology such as Trichoptera (caddis flies) and Diptera (flies).

In addition to this general aim, the thesis project has a number of major sub-themes:

1. The development of a new relational version of the existing Bugs database. The Bugs2000 system (Buckland, 2000) was constructed around a somewhat inefficient database structure that did not fully implement the concept of relational database architecture. A restructuring provided massive improvements in the efficiency of data retrieval, updating and data security aspects of Bugs. It also allowed for the development of more advanced searching, querying and reporting tools which can take advantage of the improved architecture. These improvements essentially make up the core component of the BugsCEP software package, the development of which is described in Chapter 2. The system is described in full, with examples, in Chapter 3.
2. The construction of a system for (semi-)quantitative environmental reconstruction/habitat description from fossil insect remains, providing easily interpretable, and consistently comparable graphical outputs. This is based on an ecological summary system that uses the modern ecology of the organisms as its reference/calibration data, and the statistical methods employed are transparent and relatively simple. It provides facilities for compensating for unequal sample sizes and abundances, as are common in (palaeo)ecology. Inevitably, the methods employed are based on a number of existing classification and visualisation concepts, although they are provided here for the first time connected directly to a database of Coleopteran ecology and fossil records. This system, which makes up the BugStats package component, is described in Chapter 4.
3. The implementation, and enhancement of the Mutual Climatic Range (MCR) method (Atkinson *et al.*, 1986) for deriving palaeotemperatures from fossil beetle assemblages, as a component in Bugs. MCR was previously available only as either MS-DOS based software or by somewhat laborious, and error prone, manual overlaying of transparency films. A version running in a graphical (Windows) environment, as developed here, would have been a significant improvement in itself. The aim here was to improve the availability of the method, provide improved graphical outputs, and explore the possibilities for improving the accuracy or precision of the MCR method through statistical techniques. This amounts to the BugsMCR package component, which is described in Chapter 5.
4. The testing of the thermal and environmental reconstruction software, developed in connection with this thesis, on a number of datasets, including those from published modern and fossil studies, as well as those produced by the author specifically for this thesis. These studies are presented in Chapter 6.

In addition to these specific aims, the wider Bugs project also endeavours to:

5. Make the Coleopteran fossil record of Europe publicly available through a single, downloadable source: <http://www.bugscep.com>
6. Make the process of interpreting fossil insect remains more efficient, by reduce the time necessary for looking up biology and distribution data, and performing routine data compilation/summary tasks.
7. Provide a system for the recording and storage of species list and abundance data.

These latter points have been fundamental concepts behind all previous versions of Bugs, and the latest version, BugsCEP, improves on the work of these and adds many improvements.

1.2 Scientific Background

As with any piece of science, this work is a building block in a developmental history. Although the software created here contains a number of innovations, and is the first of its kind in many respects, its development owes a lot to those who have preceded it. Aside from the numerous previous versions of Bugs (which are briefly described in Chapter 2), there are other Quaternary databases in existence that have influenced the development. The statistical methods implemented in BugsCEP (MCR, jackknifing, environmental reconstruction and coefficients of correlation) contain both original, derived and applied components.

Although the collation and storage of Quaternary entomology data was computerised relatively early (Sadler *et al.*, 1992), the development and application of quantitative methods to fossil beetle data have lagged behind some other proxy data fields. Palynologists, for example, have developed advanced numerical methods for landscape reconstruction from pollen assemblages (e.g. Sugita *et al.*, 1999). Although most authors routinely include summary statistics for numbers of beetle taxa and individuals in publications, very few attempt quantitative environmental reconstructions.

1.2.1 Databases in Quaternary science

The generally large datasets of Quaternary science make it an ideal subject for database construction. For each site within a project there may be several sampling locations (e.g. boreholes, cores, archaeological features/structures) which can result in numerous samples, for each of which there will be abundance data for any number of species (Figure 1.1). It is easy to see, then, that the individual data items can quickly amount to hundreds or thousands depending on the proxy type and preservation within the samples. Several Quaternary databases are available, perhaps the most widely used proxies being pollen and vertebrates (e.g. EPD, FAUNMAP, see section 1.2.4).

The scope for variation in abundance is enormous. This is not only a product of the natural diversity of organisms in differing environments, but a combination of this and sampling and other taphonomic factors. For example, the (early-mid Holocene) medium diversity, low abundance site of Hemavan, Sweden, analysed in this thesis (see Chapter 6) has ten samples, 61 species and 119 abundance counts. The (Lateglacial) high diversity, variable abundance site of Saint Bees in Cumbria, England, (Coope & Joachim, 1980) is made up of 35 samples, 283 species and 1 363 abundance values. These numbers are small when compared to those encountered in pollen analyses, where the microscopic nature of grains and spores and the relative ease of identification allow for much larger quantifications. The quantities, and the variations in them, are of course extremely important in the interpretation of samples, and especially when considering the relative reliability of reconstructions based on those samples. This is discussed in more detail in Chapter 4 with particular reference to quantitative methods. An enumeration of the numbers of data items in BugsCEP can be found in Chapter 3.

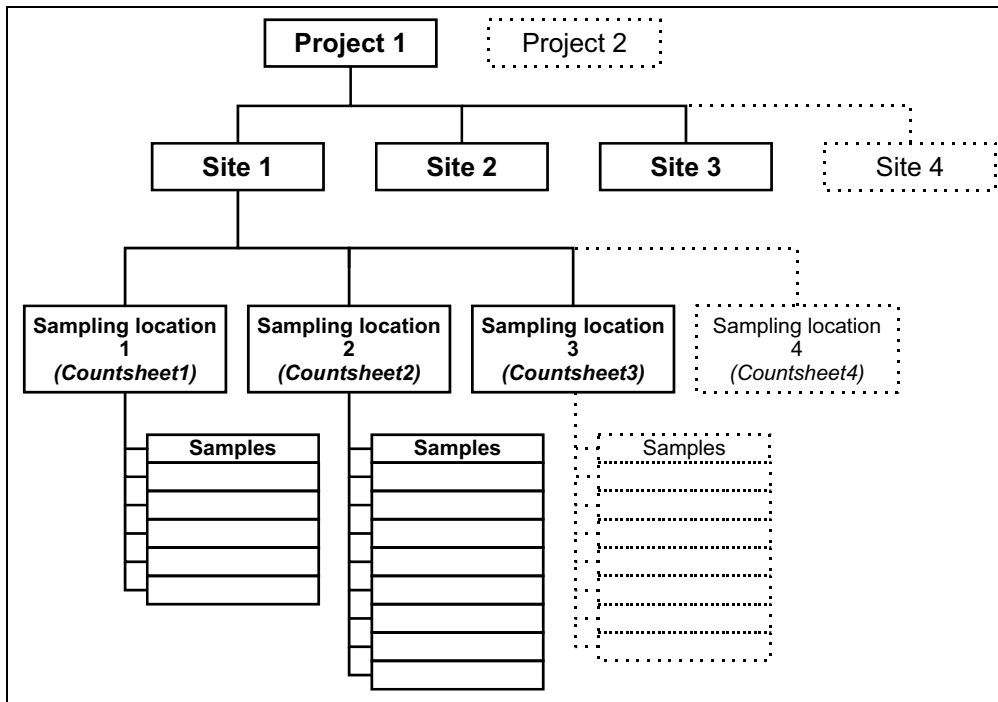


Figure 1.1. Typical Quaternary science sample hierarchy from project to sample level. *Italics* show where the BugsCEP name differs from the common usage. A sample contains abundance data for each species found in it (see Table 1.1).

1.2.1.1 General Quaternary data structure

The vast majority of Quaternary data can be displayed using a simple cross-tabulation (crosstab) of species against samples as shown in Table 1.1. Abundance counts, or the number of individuals, are recorded for the occurrence of each species in each sample. In the majority of fossil insect works, these are usually the minimum number of individuals (MNI) represented by the fossil exoskeleton parts (sclerites) found. Although the crosstab structure is an easily understandable form for humans, it is inefficient for data storage due to the potential for empty, or zero abundance cells, which create dead space in the table. In a database management system (DBMS) this generally not only leads to an increase in file sizes, but also breaks some of the guidelines for relational database structure. The implications of this for the efficiency of data retrieval are considerable, and although BugsCEP displays abundance data in crosstab form, it stores it in a more efficient manner, the mechanics of which are described in section 3.1.3.

Table 1.1. Typical layout of a Quaternary data cross-tabulation, referred to as *countsheets* in BugsCEP and this thesis.

Site Name	Sample 1	Sample 2	Sample 3	...Sample n
Species A	abundances			
Species B				
Species C				
...Species z				

1.2.2 The BugsCEP structure in brief

The BugsCEP software is more than just a database in that it has a large number of custom built interfaces for data entry, retrieval and manipulation. These interfaces, along with the code and other objects behind them, collectively make up the *application* or *program* part of the Bugs Coleopteran Ecology Package. The other part contains the actual data, and is what is commonly referred to as a

database. In fact, these two parts are separate files in the BugsCEP package, BugsCEP.mdb and bugsdata.mdb, and are often referred to as the *frontend* and *backend* respectively (see also section 2.2.2). The *frontend* is the part that the user interacts with, which itself interacts with the *backend* in order to fulfil the user's requests. Inevitably the functionality and quality of the *frontend* dictates how easily users can access the data. A good *frontend* will be transparent, in that the user will hardly be aware of the physical structure of the database, and user friendly, eliminating the need for knowledge of Structured Query Language (SQL) or general database skills.

1.2.3 What is BugsCEP and what does it replace?

Prior to the development of the initial concept of a Quaternary entomology database by Sadler *et al.* (1992), correlation of ecological, distributional and fossil data relied either upon the personal knowledge of the researcher or upon extensive literature searches, sometimes extracted onto a card index. Bugs initially supplemented, and later significantly reduced, reference literature searches by abstracting the most important bits of text describing the biology and distribution of species into a database. Thus a researcher could access a large part of the information needed to at least begin interpreting fossil beetle assemblages rapidly from a computer system, rather than having to use the library and thumb through books and articles. Although BugsCEP includes enough data to come a long way in interpretation, reference to the original papers is still recommended for detailed aspects of species ecology. In addition, for work in some countries, the geographical range of the source literature for a number of species is not yet extensive enough for the system to be completely independent of external literature.

The original concept of the database has been further developed to include climate and environment reconstruction, statistics, and comprehensive data management facilities. Some of these facilities were previously only available through the use of additional software, or through manual calculations.

1.2.4 Related databases

There is currently no existing system comparable with BugsCEP in terms of the scope its data and the features it provides. BugsCEP is not only a database of fossil sites and their abundance data, but also a database of modern biology, distribution and bibliographic data among other things (see Chapter 3). In addition, it includes a number of tools for climate and environmental reconstruction, and data management the likes of which are not found in any other single database system for any other proxy. This integration of a software package and a reference database for modern and fossil data is currently unique, but similar systems are in development (see e.g. SEAD below).

The list of existing and planned databases below is in no way exhaustive, and is only intended to direct the reader towards some important resources. Databases with limited regional and subject scope have been excluded, although there are a number of interesting databases with web interfaces available. There are also a number modern ecology and habitat databases that could be of use to Quaternary scientists, but contain no Quaternary data in themselves (e.g. EUNIS biodiversity database, Ground Beetles of Irelandⁱ). A number of the databases below, in addition to BugsCEP, are available from the NOAA Paleoclimatology website at <http://www.ncdc.noaa.gov/paleo/paleo.html>, which is part of the World Data Center Systemⁱⁱ repository for scientific data.

EDDI - European Diatom Database

<http://craticula.ncl.ac.uk/Eddi/>

A web based system including a variety of diatom datasets, training sets and transfer functions to aid in environmental reconstruction from lake deposits. Data have been collated from a large number of

ⁱ EUNIS: <http://eunis.finsiel.ro/eunis/>; Ground Beetles of Ireland: <http://www.habitas.org.uk/groundbeetles/>

ⁱⁱ <http://www.ngdc.noaa.gov/wdc/wdcmain.html>

sources, and a large number of images are included. The system also allows users to see distribution maps online, and upload data for reconstruction of pH, conductivity and other variables by a variety of methods. Software is available for download which is compatible with the online database.

EPD – European Pollen Database (and other regional equivalents)

http://www.ncdc.noaa.gov/paleo/epd/epd_main.html

Although it contains no ecological reference data, the EPD deserves a mention as one of the most widely used Quaternary databases in existence. The database provides geographical metadata, bibliographic information and raw data for a considerable number of palynological investigations throughout Europe. Related projects include equivalent data for Africa, Canada, and North America (NAPD) and South America, in addition to a Global Pollen Database (GPD). Some facility for the storage of other plant macrofossil data is built in to the structure, although the latter is under revision at the time of writing. Some of the data are available online through map (WebMapperⁱⁱⁱ) and query based search engines, although without an integrated GIS functionality. It is also possible to download the EPD as database tables, to which an SQL interface is available on request. Related projects include regionally specific climatic, landscape and ecological calibration data. A variety of external statistical and graphical tools are available for pollen data (e.g. Tilia and TiliaGraph, C2^{iv}).

ADS – Archaeological Data Service

<http://ads.ahds.ac.uk/>

This UK based large scale database project contains an enormous amount of archaeological data, but only a limited amount of environmental archaeological data. The latter datasets vary in form, and can only be queried at the metadata level. That is to say it is possible to find sites with environmental data, but individual site datasets must be extracted in order to undertake further analyses. There is great potential for the integration of environmental and archaeological datasets into the search system. No modern reference data is included, and only a limited amount of environmental proxy data is available.

FAUNMAP - late Quaternary distribution of mammal species in the United States

<http://www.ncdc.noaa.gov/paleo/fauna.html>

Funded by the US National Science Foundation, FAUNMAP includes data from about 2 919 sites spanning the last 40 000 years. The database was created to map the past distribution of mammals with an aim towards providing increased understanding of the evolution of mammalian communities. The system is accessible online, and includes search and GIS (map) interfaces for data from archaeological and palaeontological sites. The database is also intended to aid in palaeoenvironmental reconstruction, especially when combined with other geographically based proxy data. A number of international regional versions of the database exist.

SEAD – Strategic Environmental Archaeology Database

<http://www.sead.se/> (Buckland *et al.*, 2006)

Currently in the early stages of construction, SEAD is designed as a multiproxy database system for the storage and analysis of data primarily relating to environmental archaeological investigations. The system includes the ability to store large amounts of site based metadata, bibliographies and ecological reference data along with project management data. It is designed to handle a number of proxy data sources including insects, pollen, molluscs, soil properties and plant macrofossils. The initial version will be limited to Swedish data, and it will eventually be available both online and through local client interfaces. SEAD is currently under development at the Environmental Archaeology Lab, Umeå

ⁱⁱⁱ <http://www.ncdc.noaa.gov/paleo/webmapper.html>

^{iv} Tilia: <http://www.ncdc.noaa.gov/paleo/tiliafaq.html> or <http://museum.state.il.us/pub/grimm/> ;
C2: <http://www.campus.ncl.ac.uk/staff/Stephen.Juggins/software/c2home.htm> (not only pollen data)

University in Sweden. The author of this thesis is the lead developer in the SEAD project, and large parts of its interface and structure are derived from those in BugsCEP.

LNED – Late Neogene Ecosystems Database (working name)

In its early stages, LNED represents a consortium of American Quaternary scientists and environmental archaeologists working towards a unifying database of biological proxy sources. The LNED team are working with a number of the representatives of the other databases mentioned in this section to create either a single database, or series of linked databases, that would first encompass the American data, and then be expanded to include global data. In achieving this, the system will then be used as the basis for developing tools which will enable the advanced, multi-proxy querying of data in terms of climate and environmental changes over varying timescales.

1.2.5 Taxonomy and fossils

The binomial system of taxonomy, initially systematized by Carl von Linné (Linnaeus) (Knapp, 2000) provides the necessary baseline for any ecological or palaeoecological study which employs plant or animal data. Its purpose is to divide the animal and plant kingdom up into manageable units, previously by division on morphology, but more recently by genetic similarity. Morphological distinctions are extremely useful in the differentiation of fossil fragments, and a modern reference collection arranged in taxonomic order is indispensable when identifying fragments, as similar species are generally close to each other in the collection. Genetic distinctions, on the other hand, are currently of little use in palaeoecology due to the poor preservation of DNA and the time and cost that would be involved in getting DNA work done on every difficult fragment. Names have been often revised, especially since the advent of DNA techniques which have revealed numerous groupings to be incorrect, and there is thus a synonymy associated with the majority of current species names which allows for reference of the same species under various names throughout history. Statement of the taxonomic system followed in a database is therefore essential, and a list of synonyms practical, if it is to be useful internationally and in several fields. Most countries have their own taxonomic systems, which are usually similar to each other, but reflect the local faunas and the history of regional entomology to an extent. There are Internet based systems in development, such as the Global Biodiversity Information Facility (GBIF, 2006) which are aimed at uniting, or at least meshing, national systems and providing an international node for taxonomic information, including data on the availability of reference specimens in museums.

BugsCEP uses a taxonomy based upon Lucht (1987) as revised by Böhme (2005), and Gustafsson (2005) (with changes where noted), and with some revision at the family and subfamily level, according to Lawrence & Newton (1995). Taxonomic codes, a modified form of the Central European Codes of Lucht (1987), are used throughout the database to provide a unique numerical identifier to every taxon, and provide the taxonomic order.

Species are not always fully identifiable as fossils, either because of poor preservation masking characters necessary for splitting between species, or because some species simply are not identifiable to species level on the individual parts that are found fossil (most commonly the head, thorax and elytra). The latter varies between groups, with genera smaller in size being generally more difficult to identify. The skill of the individual researcher and access to modern reference collections, for comparative material, are also a factor. BugsCEP includes a number of taxon records that reflect these difficulties, and others for particularly difficult species. Almost every genus is ended with ‘sp.’ (species) and ‘spp.’ (species plural) records which can be used to record individuals identified to generic level, as can the ‘indet.’ (indeterminate) records that exist for some more difficult families. In addition, there are a number of doublets for commonly hard to split species, such as *Nebria brevicollis/salina* (F.)/Fairm. & Lab. – which indicates an individual identified as either the ground beetle *Nebria brevicollis* (F.) or *Nebria salina* Fairm. & Lab. Finally, there are particular groups (grp.) of species that are hard to resolve, such as the mould beetles group *Latridius minutus* (grp.) (L.), which includes the species *L. pseudominutus* (Strand), *L. anthracinus* (Mann.) and *L. minutus* (L.).

The interpretive implications of all of the above are not always simple, although the level of detail in environmental reconstruction is usually reduced with lower taxonomic resolutions. The size of the genera, but more importantly the amount of variation in species habits and ecological preferences within the genera, affect how useful a generic level identification is. For example, a generic identification of the water beetle *Agabus* sp. can tell us that there was probably water close by, but not whether it was likely to be running or standing water, whereas the identification of *Agabus paludosus* (F.) is almost definitely an indication of running water (Nilsson & Holmen, 1995). Interpretive aspects will be discussed further in Chapter 4, along with the use of coded habitat descriptions.

1.2.6 The archaeological and contemporary contexts

This thesis is not what would currently be called a piece of traditional archaeological work, and some archaeologists would perhaps not even class it as archaeology at all. Environmental archaeology is a broader form of the archaeology discipline that is inherently multidisciplinary, and draws on methodology from numerous fields of science. It is also largely an empirical science, its practitioners generally requiring quantitative support for statements on the nature of the past. BugsCEP is, among other things, a tool for palaeoentomology, a science which can be extremely useful in providing data on past human activity and the natural environment. Under the right conditions, insect remains can be preserved in deposits on archaeological sites and in the sediments around them. The latter sediments are frequently overlooked or omitted from archaeological investigations to cut costs, but are extremely important for providing data to help understand the nature of the interactions between the occupants of a site and the environments around them. There are a considerable amount of data from archaeological sites in BugsCEP's database (see Chapter 3), and the software has the facilities for handling dating methods more typical of archaeological excavations, such as artefact typology and period designation, as well as radiometric methods more commonly used in Quaternary geology.

An understanding of the past environmental impacts of people is extremely important for policy decisions concerning sustainable development, the past being the key to the future. Similarly, investigations into the present day effects of human populations on biodiversity should always include an awareness of the long term past. An archaeological or geological component in research lines associated with these subject areas can provide important information on the probable consequences of planned actions if they have parallels in the past. Palaeoentomology, along with other proxy methods, can help by providing data on past environmental changes, and in combination with archaeological data provide evidence of the human component in these. It is becoming more common for Quaternary research to include an archaeological component, or at least an awareness of its importance, and the construction of databases with datasets and tools common to archaeology, Quaternary geology and biology are helping to bridge the gaps between these closely related research fields (and even occasionally influencing national policies^v). The Bugs EcoCode habitat classification system (see chapters 3 and 4), which was developed as part of this thesis, for example, has a specific class, *General synanthropic*, for species known to be found in close association with humans. It also includes a number of other classes that are particularly useful in describing past onsite (archaeological) activity: *Dung/foul habitats*, *Dry dead wood* and *Ectoparasites*. The last of these are almost only ever found fossil in archaeological deposits, due to the human habit of concentrating animals into small spaces and thus increasing both the potential for parasite survival and their deposition in preserving sediments. The use of this classification system in the analysis of 'natural' deposits, in collaboration with archaeologists, may help provide information on the probability of human influence in shaping the landscapes that lead to their deposition.

As well as the inclusion of archaeological groups in the Bugs EcoCode classification system, there are a large number of indirect relationships between the subject of this thesis and archaeology. Quaternary science and archaeology should be, and in fact have previously been intimately related. Many of the methods used in archaeology, such as dendrochronology, pollen analysis, radiocarbon dating and even the logic of stratigraphy were developed in Quaternary geology, and archaeologists would be hard

^v As the sudden government awareness of the importance of climate change has demonstrated (e.g. Stern, 2006).

pressed to achieve their goals without them. Although much archaeological work is site based, focussing on the deposits left behind by the people and animals that occupied a specific place, the data obtainable from archaeological sites have only a limited potential without knowledge of the surrounding environment, and the interaction of people with it. Activities from simply the gathering of firewood, to deforestation and farming all leave their signals in sediments in different ways, and at different scales. It is difficult to understand the past course of human activities at a site, and their impact on the surroundings, without an insight into the background conditions there. To understand these things we need empirical data from proxy sources – we cannot measure them directly, so we must rely on the examination of things that were there at the time to tell us about the concurrent conditions. It is simply not possible to stand in a present day landscape, and, by a process of imagination or projection, however ‘soundly based in theory’ it may be, obtain a reasonably accurate picture of any distant past form of that landscape.

To summarize, empirical data on past environments are essential for interpreting and understanding the history of human-nature interactions. Fossil beetles can provide us with excellent proxy data for these interactions, and BugsCEP is a tool that can help in the manipulation and interpretation of these data.