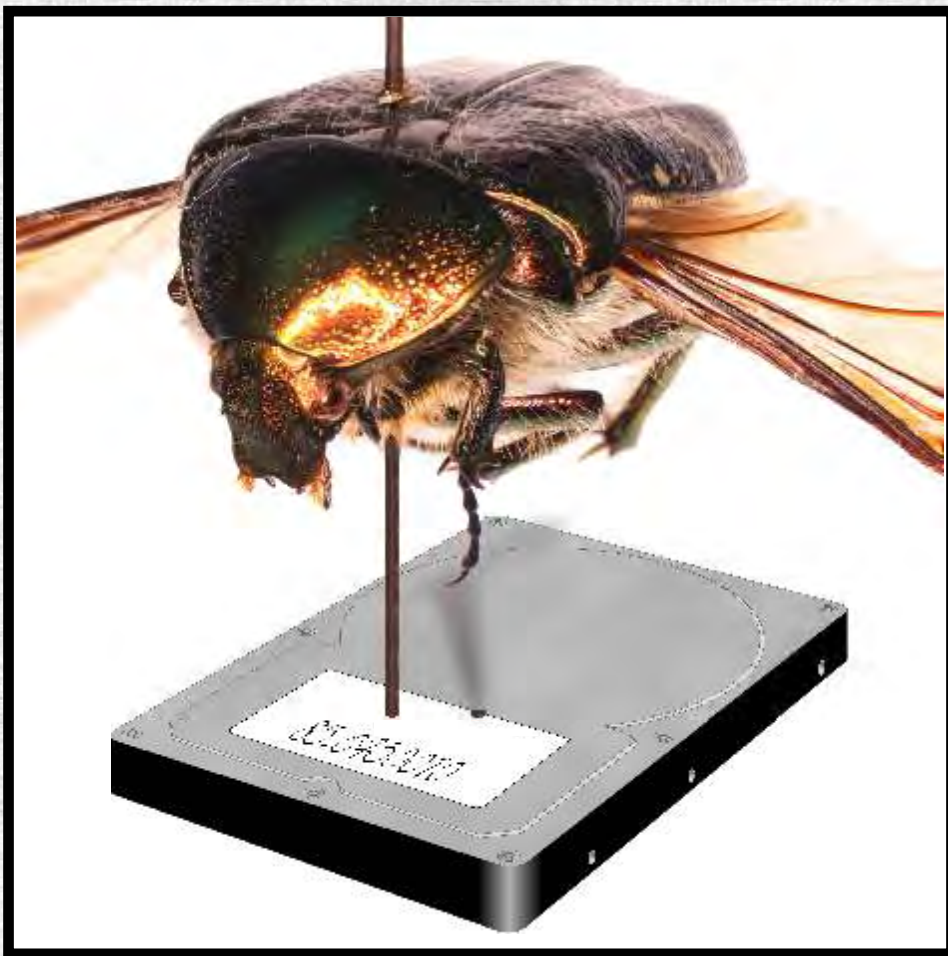


**The Development and Implementation of Software for
Palaeoenvironmental and Palaeoclimatological Research:
The Bugs Coleopteran Ecology Package (BugsCEP)**

Philip I Buckland



Umeå University
Department of Archaeology
& Sámi Studies

Archaeology and Environment 23

**The Development and Implementation of Software for
Palaeoenvironmental and Palaeoclimatological Research:
The Bugs Coleopteran Ecology Package
(BugsCEP)**

A dissertation for the degree of Doctor of Philosophy
Umeå University 2007
Department of Archaeology and Sámi Studies
SE-901 87 Umeå

Author

Philip I Buckland

Title

The Development and Implementation of Software for Palaeoenvironmental and Palaeoclimatological Research: The Bugs Coleopteran Ecology Package (BugsCEP)

Abstract

This thesis documents the development and application of a unique database orientated software package, BugsCEP, for environmental and climatic reconstruction from fossil beetle (Coleoptera) assemblages. The software tools are described, and the incorporated statistical methods discussed and evaluated with respect to both published modern and fossil data, as well as the author's own investigations.

BugsCEP consists of a reference database of ecology and distribution data for over 5 800 taxa, and includes temperature tolerance data for 436 species. It also contains abundance and summary data for almost 700 sites - the majority of the known Quaternary fossil coleopteran record of Europe. Sample based dating evidence is stored for a large number of these sites, and the data are supported by a bibliography of over 3 300 sources. Through the use of built in statistical methods, employing a specially developed habitat classification system (Bugs EcoCodes), semi-quantitative environmental reconstructions can be undertaken, and output graphically, to aid in the interpretation of sites. A number of built in searching and reporting functions also increase the efficiency with which analyses can be undertaken, including the facility to list the fossil record of species found by searching the ecology and distribution data. The existing Mutual Climatic Range (MCR) climate reconstruction method is implemented and improved upon in BugsCEP, as BugsMCR, which includes predictive modelling and the output of graphs and climate space maps.

The evaluation of the software demonstrates good performance when compared to existing interpretations. The standardization method employed in habitat reconstructions, designed to enable the inter-comparison of samples and sites without the interference of differing numbers of species and individuals, also appears to be robust and effective. Quantitative climate reconstructions can be easily undertaken from within the software, as well as an amount of predictive modelling. The use of jackknifing variants as an aid to the interpretation of climate reconstructions is discussed, and suggested as a potential indicator of reliability. The combination of the BugStats statistical system with an enhanced MCR facility could be extremely useful in increasing our understanding of not only past environmental and climate change, but also the biogeography and ecology of insect populations in general.

BugsCEP is the only available software package integrating modern and fossil coleopteran data, and the included reconstruction and analysis tools provide a powerful resource for research and teaching in palaeo-environmental science. The use of modern reference data also makes the package potentially useful in the study of present day insect faunas, and the effects of climate and environmental change on their distributions. The reconstruction methods could thus be inverted, and used as predictive tools in the study of biodiversity and the implications of sustainable development policies on present day habitats.

Keywords: environmental archaeology, Quaternary science, Coleoptera, beetles, database, environmental reconstructions, climate reconstructions, software, Mutual Climatic Range, MCR, palaeoentomology

Philip I Buckland, Dept. Archaeology and Sámi Studies, Umeå University, SE-901 87 Umeå, Sweden

Archaeology and Environment 23

**The Development and Implementation of Software for
Palaeoenvironmental and Palaeoclimatological Research:
The Bugs Coleopteran Ecology Package
(BugsCEP)**

Philip I Buckland

Environmental Archaeology Lab.
Department of Archaeology and Sámi Studies
2007



This work was made possible by financial support from the Bank of Sweden Tercentenary Foundation.

Illustrations by the author unless otherwise stated.

Cover: *Cetonia aurata* (L.), the rose chafer (gräsgrön guldbagge), impaled on a hard disk drive. The beetle is a composite Extended Depth of Field (EDF) image made from 15 microscope photographs, using Nikon's EclipseNet software.

© Philip I Buckland

Printed in Umeå by Solfjädern Offset AB.
ISBN 978-91-7264-298-0
ISSN 0281-5877

Contents

List of figures	v
List of tables	ix
List of abbreviations	x
Database terminology	xi
General terminology	xi
The structure of the thesis.....	xii
Acknowledgements	xii
1 Introduction	1
1.1 Aims of the Thesis	1
1.2 Scientific Background.....	2
1.2.1 Databases in Quaternary science.....	2
1.2.1.1 General Quaternary data structure.....	3
1.2.2 The BugsCEP structure in brief	3
1.2.3 What is BugsCEP and what does it replace?.....	4
1.2.4 Related databases	4
1.2.5 Taxonomy and fossils	6
1.2.6 The archaeological and contemporary contexts	7
2 The Development of BugsCEP	9
2.1 Introduction.....	9
2.2 Database and Software Background.....	9
2.2.1 Relational database design	9
2.2.2 The BugsCEP database structure	11
2.2.3 Overview of BugsCEP software features.....	13
2.3 Developmental Strategy.....	14
2.3.1 Primary developmental aims.....	15
2.3.2 Secondary aims	15
2.3.3 Development platform	16
2.4 Database Structural Changes	16
2.4.1 Bugs database structure and contents.....	16
2.4.2 Conversion of Bugs2000 data	18
2.4.3 Database structural compromises.....	18
2.4.4 Problems and potential problems with the BugsCEP structure.....	20
2.5 Program Development	22
2.5.1 Summary of deficiencies in the previous version (Bugs2000).....	22
2.5.2 Main improvements in the BugsCEP system	22
2.6 Testing	24
2.6.1 Developer testing	24
2.6.2 User-based testing.....	25
2.7 Presentations and Publications.....	25
2.7.1 Publications about or with direct use of Bugs.....	25
2.7.2 Unpublished presentations	26
2.7.3 Poster presentations.....	26
2.7.4 A note on other publications using Bugs (2000/CEP).....	26
2.8 A Brief Developmental History of Bugs	27
2.8.1 Design criteria for Bugs	27
2.8.2 Implementing the design criteria.....	27
2.8.3 Version history	28
2.8.4 Bugs spin-offs	30
2.8.4.1 Slugs – Molluscan Database	30
2.8.4.2 EgBugs – The Egyptian Coleoptera Database	30
2.8.4.3 Other embryonic versions	31
2.8.5 Other Related Software	32
2.8.5.1 MS-DOS MCR Software – RECON & RECON2 and related programs.....	32
2.8.5.2 BugsMCR (standalone version)	32
2.8.5.3 BugStats (standalone version).....	32
3 BugsCEP Database System Description.....	33
3.1 Data Area Descriptions.....	33

3.1.1	Common data area.....	34
3.1.1.1	Master species list and taxonomic code	34
3.1.1.2	Synonyms	34
3.1.1.3	Taxonomic notes	35
3.1.1.4	Measurable attributes	35
3.1.1.5	Identification keys (provided by Peter Skidmore).....	35
3.1.1.6	Bibliography.....	35
3.1.2	Modern reference/calibration data.....	36
3.1.2.1	Biology and distribution text abstracts	36
3.1.2.2	Ecology summary codes (Bugs EcoCodes & Koch ecology codes).....	37
3.1.2.3	Red Data Book (RDB) classifications	37
3.1.2.4	Species associations	38
3.1.2.5	Seasons of adult activity	38
3.1.2.6	Climate calibration data (thermal envelopes and summaries)	38
3.1.3	Fossil/site data area	39
3.1.3.1	Sites	39
3.1.3.2	Site descriptive data and metadata	40
3.1.3.3	Countsheets: species lists, samples & abundance data	40
3.1.3.4	Sample dates.....	41
3.2	Record Enumeration & Database Size	42
3.2.1	Sites overview	44
3.2.2	Countsheets: species lists, samples & abundance data	45
3.2.3	Sample dates.....	46
3.2.3.1	Radiometric dates	47
3.2.3.2	Calendar dates	48
3.2.3.3	Period dates	49
3.3	User Base	50
3.4	The BugsCEP Program - Detailed Description.....	51
3.4.1	BugsCEP Main Screen – basic data retrieval	52
3.4.1.1	Section (1): Title bar, menu bar and component buttons.....	54
3.4.1.2	Section (2): Navigation panel.....	54
3.4.1.3	Section (3): Information tabs, and section (4) Information area.....	56
3.4.1.4	Section (5): Additional buttons and administrative controls	59
3.4.2	Site and Countsheet Management – adding/retrieving sites and abundance data.....	62
3.4.2.1	The Site Manager	62
3.4.2.2	Site Information screen	63
3.4.2.3	Managing countsheets and entering abundance data.....	64
3.4.2.4	Importing MS Excel spreadsheets	67
3.4.2.5	The Date Explorer	70
3.4.3	Environmental reconstruction - BugStats.....	71
3.4.3.1	Calculations options	73
3.4.3.2	Chart options	74
3.4.3.3	Creating EcoCode outputs.....	74
3.4.3.4	Seeing the Bugs EcoCode definitions	76
3.4.3.5	Calculating coefficients of similarity	76
3.4.4	Climate reconstruction - BugsMCR	78
3.4.5	The Search Interface.....	83
3.4.5.1	A worked example.....	85
3.4.6	The help files.....	87
3.5	Reporting & Exporting Functions.....	88
3.5.1	Single species data	88
3.5.2	Bibliographic data	89
3.5.3	Site reports.....	89
3.5.4	Reporting search results	90
3.6	The Addition of Taxa, Biology/Distribution Data and References.....	93
4	BugStats: Software for Environmental Reconstruction and Statistics from Beetle Assemblages.....	95
4.1	Why BugStats?.....	95
4.2	Background	96
4.2.1	Quantitative methods in Quaternary science	96

4.2.1.1	Statistics in (palaeo)entomology and environmental science.....	96
4.2.1.2	Taphonomy and the representation of taxa in samples	97
4.2.2	Biogeography, evolution and palaeoecology	97
4.2.3	Classification.....	98
4.2.3.1	Habitat classification and palaeoecology	100
4.2.3.2	Further notes on habitat classification methods	101
4.3	The Bugs EcoCode Classification System, and the BugStats Environmental Reconstruction Software	101
4.3.1	Bugs EcoCode classification system description	101
4.3.2	Bugs EcoCode designations and their implications	104
4.3.3	EcoFig calculations, transformation and standardization.....	105
4.3.3.1	Excluding taxa not identified to species level	106
4.3.3.2	Logarithmic transformation	106
4.3.3.3	Standardization.....	109
4.3.3.4	EcoFig diagram creation	110
4.3.3.5	Sample by sample EcoCode report	111
4.3.3.6	Alternative standardization possibilities	111
4.3.3.7	Final note on the use of BugStats options	112
4.3.4	Known issues with EcoCodes	112
4.3.4.1	Geographical variation in habitat preference	112
4.3.4.2	Indicators and standardization.....	113
4.3.4.3	Indicators and diversity	113
4.3.5	Correlation coefficients.....	113
4.4	The mechanics of BugStats.....	114
4.4.1	EcoFig calculations and diagram creation.....	114
4.4.2	EcoCode reports	115
4.4.3	Coefficient calculation	115
4.5	Further developments and additional methods	115
4.6	Conclusions.....	117
5	BugsMCR: Software for MCR Temperature Reconstruction from Beetle Assemblages....	119
5.1	Background and Software Development	119
5.1.1	The mutual climatic range (MCR) method in brief.....	119
5.1.2	The BugsMCR implementation	121
5.1.2.1	Species thermal envelopes	122
5.1.2.2	Calculations and overlaps	124
5.1.2.3	Graphs	126
5.1.2.4	Advanced MCR.....	126
5.1.3	Predicting potential changes in geographical range	127
5.2	MCR – Problems and Possibilities	130
5.2.1	Requirements for quantitative reconstructions.....	130
5.2.2	MCR and urban deposits.....	132
5.2.3	Problems with MCR correction/calibration	132
5.2.4	Ubiquity analysis.....	134
5.2.5	Calculation of relative warm/cold components.....	134
5.2.6	The use of jackknifing to investigate or enhance the reliability of results	135
5.2.6.1	Multiple removal jackknifing as a reliability index	139
5.2.6.2	Potential problems with jackknifing and MCR	139
5.3	Conclusions and Future Directions.....	140
6	Case Studies – testing BugsCEP with real data	141
6.1	Modern Case Study: Forest-Farmland Pitfall Trap Transects in Lammi, Finland.....	142
6.1.1	Aims.....	142
6.1.2	Introduction.....	142
6.1.3	Methods.....	143
6.1.4	Results and preliminary discussion	144
6.1.5	Further discussion	148
6.1.6	Conclusions and implications for BugStats.....	149
6.2	Fossil Case Study: 140 000 year Peat Sequence, La Grande Pile, France.....	150
6.2.1	Aims.....	150
6.2.2	Methods.....	150

6.2.3	Results and general comparison	151
6.2.4	Discussion and detailed comparison of specific habitat groups	155
6.2.5	How are Ponel's faunal and chronological groups affected by the use of standardized results and sample correlations?	160
6.2.6	Conclusions	164
6.3	Fossil Case Study: Early Holocene Environmental and Climate Change at Hemavan, Northern Sweden.....	164
6.3.1	Aims	164
6.3.2	Site, samples and methods.....	165
6.3.3	Results	168
6.3.4	Discussion	172
6.3.5	Evidence for early Holocene climate change	173
6.3.6	Conclusions	174
6.4	Fossil Case Study: Two 'Wells' at the Archaeological Site Lockarp 7B, Sweden.....	174
6.4.1	Aims and introduction	174
6.4.2	Methods.....	175
6.4.3	Results and discussion.....	176
6.4.3.1	Feature 14495.....	179
6.4.3.2	Feature 26551.....	180
6.4.4	Conclusions	181
6.5	Preliminary Fossil Results from Lake Njulla, Abisko, Sweden.....	182
6.5.1	Introduction	182
6.5.2	Preliminary results.....	183
6.5.3	Conclusions	184
6.6	Modern Case Study: Pitfall Trap Data from the area of Gården under Sandet (GUS), Greenland.....	184
6.6.1	Aims and introduction	184
6.6.2	Results	184
6.6.3	Discussion	185
6.6.4	Conclusions	187
6.7	Fossil Case Study: Climate and Environmental Change in Europe over the Past 20 000 ¹⁴ C years Reconstructed from Coleopteran Remains	187
6.7.1	Aims	187
6.7.2	Introduction	188
6.7.3	Sites	188
6.7.4	Methods.....	192
6.7.5	Results and discussion.....	193
6.7.6	An experiment in large scale environmental reconstruction, and a test of the BugStats %SumRep standardization method.....	201
6.7.6.1	Relationships between reconstructed temperatures and habitats	204
6.7.7	Conclusions	205
7	Conclusions.....	207
7.1	Wider Applications	208
7.2	Databasing the Humanities	208
7.3	Future Directions and Final Thoughts.....	209
8	References.....	211

List of figures

Figure 1.1. Typical Quaternary science sample hierarchy from project to sample level. <i>Italics</i> 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).	3
Figure 2.1. Relational database terminology, and the normalized taxonomic index structure of the SEAD database (Buckland <i>et al.</i> , 2006).	10
Figure 2.2. BugsCEP program frontend-backend structure, showing user, developer and data manager interaction.....	12
Figure 2.3. BugsCEP backend (database) structure. Boxes represent data tables with lists of their fields enclosed. Lines represent relationships between the tables (see Figure 2.1 for explanation). Note that a number of lookup and reference tables have been omitted to improve clarity. The three table groups, or data areas, Common, Modern and Fossil/Site are explained in Chapter 3.....	13
Figure 2.4. Bugs2000 database structure. Note that the relationships between the Biblio and SHNEW are symbolic, and represent programmed links rather than enforced relationships.....	17
Figure 2.5. BugsCEP database structure, repeated from Figure 2.3 to allow for easier comparison with the structure of the previous version, Bugs2000 (Figure 2.4).	17
Figure 2.6. Start-up screen of Slugs v.2 prototype molluscan database	30
Figure 2.7. Start-up screen of EgBugs – the Egyptian Entomology Database.	31
Figure 3.1. Screenshot: BugsCEP start-up screen.	33
Figure 3.2. Structure of the Red Data Book (RDB) system and code (classification) data area.....	38
Figure 3.3. Structure of the site and samples area of the database, illustrating how data items are displayed (top) and saved (bottom). Note that the empty cells in the countsheet are not saved in the <i>TFossil</i> table.	41
Figure 3.4. Bar graph showing number of sites per country. Lighter parts of the bars represent sites with only summary information and no abundance data. ‘Other’ shows the total sum for all countries with less than nine sites per country.	44
Figure 3.5. Geographical location of sites in the BugsCEP database. Mercator projection.	45
Figure 3.6. Map showing location of sites with stratigraphic sequences, archaeological contexts and no countsheets in the BugsCEP core region.....	46
Figure 3.7. Histogram showing number of radiometric dates per 5 000 year period (bin) over the last 100 000 radiometric years.....	47
Figure 3.8. Histogram showing number of radiometrically dates per 1 000 year period (bin) over the last 20 000 radiometric years.....	48
Figure 3.9. Histogram showing number of calendar dates per 250 year period over the last 5 000 years (corrected to calendar BP where present = 1950 for limited comparability with radiometric ages).	48
Figure 3.10. Bar chart showing number of dates stored per time period. The ‘Other’ category is the sum of all periods with less than ten dates.	49
Figure 3.11. Screenshot: BugsCEP main screen showing biology and distribution data for <i>Cicindela sylvatica</i> L.	53
Figure 3.12. Screenshot: Bibliography pop-up showing references for <i>Cicindela sylvatica</i> L.....	53
Figure 3.13. BugsCEP Main Screen areas: (1) Title bar, menu bar and component buttons; (2) Navigation panel; (3) Information tabs; (4) Information area; (5) Additional buttons and administrative controls. (The background image is the same as Figure 3.11).....	54
Figure 3.14. Screenshot: Taxonomic Explorer (Main Screen Navigation).....	55
Figure 3.15. Screenshot: Taxonomic code browser (Main Screen Navigation).	56
Figure 3.16. Screenshot: Main screen, synonyms panel showing data for <i>Nebria rufescens</i> (Ström.).....	57
Figure 3.17. Screenshot: Fossil/Site Records panel, showing sites with radiometrically dated samples containing <i>Nebria rufescens</i> (Ström.).	58
Figure 3.18. Screenshot: Ecological Summary panel for <i>Aphodius foetens</i> (F.).	59
Figure 3.19. Screenshot: Synonym Browser, showing search for ‘ <i>gyllenhalii</i> ’. The Find dialog has been positioned to overlap the browser for ease of use.	60
Figure 3.20. Screenshot: First two couplets of the key to <i>Agriotes</i> species.....	61
Figure 3.21. Screenshot: Species associations for <i>Dyschirius nitidus</i> (Dej.).....	61
Figure 3.22. Screenshot: BugsCEP Site Manager, with the site ‘Brigg’ selected.	63
Figure 3.23. Screenshot: Site Information Screen showing details for Brigg (Buckland, 1981).	64
Figure 3.24. Screenshot: The Countsheet Manager showing the two countsheets stored for Brigg.	64
Figure 3.25. Screenshot: The Countsheet Explorer spreadsheet display, showing part of the abundance data for the countsheet ‘Brigg Column_bugsdata.xls’. Taxa names fill the first column and sample names form the remaining column headers.	65
Figure 3.26. Screenshot: Species List Editor.....	66
Figure 3.27. Screenshot: Samples Manager showing the first three samples of the stratigraphic sequence from the countsheet ‘Brigg Column_bugsdata.xls’. Sample depths have been entered for the first two samples....	67

Figure 3.28. Spreadsheet format for import of MS Excel files into BugsCEP. Note that the divider between Genus and species could also be a space.	68
Figure 3.29. Screenshot: BugsCEP Import Wizard showing stages one and two completed in the import of the hemavan_bugsdata.xls MS Excel file.	68
Figure 3.30. Screenshot: BugsCEP File Converter, having attempted to automatically convert the file hemavan_old_format.xls. Blank cells in the first few columns of the New Countsheet (bottom) indicate a failed match – the original name being shown in column ‘F1’. Names marked with an asterisk are uncertain matches. Column widths have been adjusted for clarity, by drag the right hand edge of the column headers.	69
Figure 3.31. Screenshot: Message displayed on completion of the automatic phase of imported file conversion. All problems indicated here must be corrected manually in the File Converter before import can continue.	70
Figure 3.32. Screenshot: Date Explorer showing calendar dates for first sample in the countsheet Sandnes Felt I, Greenland.	71
Figure 3.33. Screenshot: Date Explorer showing a radiocarbon dated sample from Messingham, UK.	71
Figure 3.34. Screenshot: BugStats main screen with the site Stóraborg and countsheet ‘Storaborg_bugsdata.xls’ selected. The ‘Graph title’ has been entered manually.	72
Figure 3.35. Screenshot: the first few charts of a Bugs EcoFig for Stóraborg, as opened in MS Excel for the first time.	75
Figure 3.36. Screenshot: Start of report showing a sample by sample breakdown of the taxa found at Stóraborg, including their abundances and Bugs EcoCode designations.	76
Figure 3.37. Screenshot: The coefficients calculation interface, with Stóraborg selected.	77
Figure 3.38. Screenshot: Coefficients output opened in MS Excel for first time.	78
Figure 3.39. Screenshot: The BugsMCR interface with the Saint Bees site and ‘Saint Bees Coope_bugsdata.XLS’ countsheet activated.	79
Figure 3.40. Thermal reconstruction from the Saint Bees site, using the ‘Closest to 100%’ overlap option. Note that the temperature axes may differ on output, and must be scaled manually. All results are in degrees Celcius, and the warming of both summer and winter temperatures through the sequence can clearly be seen from left to right.	80
Figure 3.41. Climate space maps for two samples from Saint Bees, as exported by BugsMCR, showing (a) a 100 % overlap scenario and (b) a non 100 % overlap scenario with a complex area of maximum overlap. See Chapter 5 for further a more detailed explanation of envelopes and overlaps.	82
Figure 3.42. Screenshot: MCR calculation progress indicator, showing that sample 35 out of 35 is being calculated.	82
Figure 3.43. Bugs Search Explorer, with interface parts as follows: (1) Criteria type tabs (top) and search criteria selection for the active tab; (2) Action buttons; (3) Current species list (search results); (4) Current search session log; (5) Report/export buttons.	83
Figure 3.44. Flow diagram illustrating the sequence of events when using the Bugs Search Explorer.	85
Figure 3.45. Search results report showing ‘Just the names’ of the taxa which are RDB classified as extinct in the UK, classified as aquatic and known from Sweden.	86
Figure 3.46. Final page of the ‘Just the names’ search results report, showing the sequence of events that lead to the final species list (the search log).	87
Figure 3.47. Screenshot: The BugsCEP help contents page, the links below the title bar providing navigation.	87
Figure 3.48. Screenshot: The report toolbar, displayed at the top of all report previews.	88
Figure 3.49. Screenshot: The Report Generator, which allows the user to specify which type of report to create. It is used in several areas of the program, and has a number of variations as shown by a and b	89
Figure 3.50. Site report, full version, showing part of the first and last pages of the 82 page report generated for the Saint Bees site. The first page shows site summary information and the first taxa from the site, and the last page shows the last few references cited in the biology and distribution data.	90
Figure 3.51. Search results report using the ‘Names; EcoCodes; RDB’ option. Each taxon name is followed by its BugsCEP habitat and Red Data Book classifications.	91
Figure 3.52. Search results report using the ‘Names; Synonyms; Taxonomic Notes’ option.	91
Figure 3.53. Search results report using the ‘Names; Biology & Distribution texts with full references’ option.	92
Figure 3.54. Report showing the first few sites from which the seven species shown in Figure 3.45, which were retrieved through the Search Explorer, are known.	93
Figure 4.1. Number of Bugs EcoCodes per taxon.	105
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.	105
Figure 4.3. <i>Unmodified</i> 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).	107

Figure 4.4. Flow diagram illustrating the sequence of events that create an EcoFig diagram from site data.	108
Figure 5.1. An illustration of the derivation of temperature values (TValues) using the MCR method, showing the mutual climatic range (MCR) for three species in climate space.	120
Figure 5.2. Thermal envelope for <i>Carabus problematicus</i> Hbst. as stored in BugsCEP, showing how taxonomic code links the envelope data to ecology data through the master species index. The original envelope, as stored in the beetle.dat data file for the MCRBirm RECON software is shown in the inset to the bottom left.	123
Figure 5.3. Screenshot: Comparison of BugsCEP taxa with names of taxa as stored in the original MCRBirm software. The 'RECON Nr' field shows the numbers used as input strings in the original RECON(struction) component of the MCRBirm software.	124
Figure 5.4. Comparison of exported climate space maps for Saint Bees (Coope & Joachim, 1980) sample s50 as processed by (a) BugsMCR and (b) RECON. Note the difference in reconstructed temperatures caused by the definition of maximum overlap cells. Both maps have been shaded, had scales added and the area of maximum overlap highlighted. The BugsMCR output has been rounded to the nearest percent for clarity, and the RECON values are in percent/10 as output by the program. The scales indicate the upper boundary of climate cells in degrees Celsius.	125
Figure 5.5. Screenshot: Advanced MCR interface, showing the sample selection panel on the left and numerous options on the right.	126
Figure 5.6. Screenshot: The BugsMCR Predictions interface with the thermal envelope limits for <i>Diacheila arctica</i> (Gyll.) selected and copied to the 'Ranges' panel.	128
Figure 5.7. Cross section of a sample climate space map, along the TRange = 21°C line, showing % of species in each TMax 1°C climate cell. Note the square nature of the curve, and the plateau like nature of the 100 % area, indicating the equal probability of any cell in this area.	133
Figure 5.8. Chart showing jackknife limits for all temperature variables for Saint Bees sample s50. The thick bars show the standard MCR results, and the extensions the jackknife extremes. The source data is presented in Table 5.6.	137
Figure 6.1. Map of case study sites discussed in this chapter. See BugsCEP for the location of other sites mentioned.	141
Figure 6.2. Mean adjusted catch (and SE) of carabids of three habitat groups along forest - farmland transects. Black/white shading represents different trapping episodes. Note that the Forest species y-axis has been rescaled in this thesis to match the other habitat scales. [Reproduced with permission from Koivula <i>et al.</i> (2004)].	142
Figure 6.3. BugStats EcoFig output (edited) for Koivula <i>et al.</i> (2004) reconstructed grouped trap data. See text for discussion of variations. Note that the x-axes are different for each raw diagram. NSpec = Number of species.	146
Figure 6.4. All taxa EcoFigs for La Grande Pile. Diagrams (a) and (b) show abundance weighted, and taxa only, standardized values respectively. Diagrams (c) and (d) show abundance weighted, and taxa only, raw values respectively. Number of taxa and abundance totals are shown in Figure 6.6.	153
Figure 6.5. Species identifications only EcoFig for La Grande Pile. Diagrams (a) and (b) show abundance weighted, and taxa only, standardized values respectively. Diagrams (c) and (d) show abundance weighted, and taxa only, raw values respectively. Number of taxa and abundance totals are shown in Figure 6.6.	154
Figure 6.6. La Grande Pile, comparison of number of taxa and individuals, showing all taxa and species level only identifications for comparison. Poneil's (a) samples, (b) faunal units, and (c) the pollen chronozones of Beaulieu & Reille (1992) are shown. Note the different scales on abundance and NSpec graphs.	155
Figure 6.7. Comparison of woodland habitat numbers for Poneil's classes: Conifer dependant, Deciduous dependant, Tree-dependent; and the equivalent raw, species level identification only, representation counts in BugsCEP: <i>Indicators: Coniferous, Indicators: Deciduous, Wood and trees</i> respectively. T = No. of Taxa, I = No. of Individuals. Sample 29 is absent from the original figure and therefore also the BugStats charts.	156
Figure 6.8. Comparison of aquatic habitat numbers for Poneil's classes: Aquatics, Standing-water, Running-water; and the equivalent for all taxa in BugsCEP: <i>Aquatics, Indicators: Standing water, Indicators: Running water</i> respectively.	157
Figure 6.9. Comparison of dung habitat numbers for Poneil's coprophagous class and the equivalents in BugsCEP: <i>Indicators: Dung, Dung/foul habitats and Pasture/Dung</i> . Note the different scales.	158
Figure 6.10. A comparison of the BugsMCR temperature reconstruction for La Grande Pile with that provided by Poneil. Numbers of taxa and individuals used in the BugsMCR calculations are provided on the right, and samples where the area of maximum overlap contains <100 % of taxa are indicated by lighter shading.	160
Figure 6.11. Comparison of Poneil's faunal units with those derived from correlation coefficients for neighbouring samples, and possible implications of BugStats standardized outputs on the original faunal	

units. Coefficient based sample groups (B1-B7.4) are shaded by the average of the coefficient values for samples in the group. 0 = no similarity, 1 = total similarity.	162
Figure 6.12. Panorama photograph of the Hemavan bog taken from the sampling location.	164
Figure 6.13. Map showing the Hemavan site and location of the sample profile.	165
Figure 6.14. Photograph of the sampling location, with remains of Engelmark (1996) section in the foreground on the left. Paul Buckland is documenting the vegetation (and simuliids) in the background.	166
Figure 6.15. Photograph showing the sampling pit at two stages of excavation. The scale is approximate and for illustration only. A dark, more humified band can be seen at c. 100 cm, and the peat-clay transition is visible at about 135 cm.	167
Figure 6.16. Bugs EcoFigs and other biological proxies for Hemavan. Standardized beetle reconstructions with (a) abundance weighting and (b) taxa only. Diagram (c) shows summary pollen percentage data, and (d) raw plant macrofossil counts. Radiocarbon dates are uncalibrated, and transferred by stratigraphic correlation.	171
Figure 6.17. MCR reconstruction of palaeotemperatures from Hemavan. Uncalibrated ^{14}C dates are shown on the diagram; 1σ calibrated ranges U-2692: 5122-6184 BP and U-2695: 8599-9243 BP (Oxcal version 4, Bronk Ramsey, 1995). Present day temperatures: TMax: 8 to 10°C ; TMin: -15 to -12°C (SMHI, 2005)	172
Figure 6.18. Stratigraphy and sample locations in feature 14495, Lockarp 7B. Stratigraphy provided by Johan Linderholm; archaeological sediment descriptions can be found in Eliasson & Kishonti, (2003).	175
Figure 6.19. Stratigraphy of feature 26551, Lockarp 7B. Insects were only preserved in the bottom 20 cm of the sequence. Stratigraphy provided by Johan Linderholm; archaeological sediment descriptions can be found in Eliasson & Kishonti, (2003).	176
Figure 6.20. BugStats output for Lockarp 7B feature 14495. Both diagrams are standardized, showing (a) abundance weighted and (b) taxa only results.	179
Figure 6.21. BugStats output for Lockarp 7B feature 26551, illustrating the importance of sample standardization (see text). Diagrams (a) and (a) show standardized abundance weighted and taxa only results respectively; diagrams (c) and (d) show raw counts, abundance weighted and taxa only respectively.	181
Figure 6.22. Correlation of cores from Lake Njulla, Abisko, Sweden. Lines indicate correlation horizons between the cores. Core group 3 was taken with a larger corer bore and was not able to penetrate as deep as the others. Core group 1 is considered as the most reliable set of cores.	183
Figure 6.23. BugStats output for GUS modern data, showing (a) standardized, taxa only reconstruction, and (b) abundance weighted, raw counts based reconstruction.	186
Figure 6.24. First two components of PCA on the GUS modern pitfall trap data, showing species names.	186
Figure 6.25. First two components of PCA on the GUS modern pitfall trap data, showing vegetation/sampling zones.	187
Figure 6.26. Map of sites used in the 20 000 ^{14}C year reconstruction, see Table 6.15 for site names and time slices. The positions of the numbers in the call-out bubbles show the approximate position of the respective site symbols (+) on the map.	189
Figure 6.27. Plot of latitudes of sites in the 20 000 ^{14}C year reconstruction, per time slice.	190
Figure 6.28. MCR temperature reconstruction for the last 20 000 ^{14}C in 1 000 year slices using all European ^{14}C dated stratigraphic sequence samples in BugsCEP (Greenland excluded). (a) Bars represent the possible range of temperatures as indicated by the beetle faunas, for TMax and TMin; (b) the percentage of species in the area of maximum overlap from which the MCR values are calculated; (c) number of taxa (thick black bars), and individuals (thin grey bars) used in the reconstruction.	194
Figure 6.29. Jackknife results from 20ka temperature reconstruction, showing (a) the maximum (black extension bars) enlargement, and reduction (white boxes) of combined envelopes caused by the removal of any one taxa. The standard MCR values are shown as in Figure 6.28a. (b) shows the percentage of taxa whose removal causes a change in the reconstructed temperature range.	196
Figure 6.30. Climate space maps for time slice 14-15K, showing percentage overlaps for (a) all taxa, and (b) with <i>Boreaphilus henningianus</i> Sahl. excluded, the thermal envelope of which is outlined and shown in bold. The cells with maximum overlap, which make up the mutual climatic range are shown with white text on black cells. Note how the removal of <i>B. henningianus</i> creates a discontinuous area of maximum overlap, the outer limits of which give the MCR temperature values. This illustrates a potentially split fauna.	197
Figure 6.31. Graphs showing relationships, in the 20 ka dataset, between (a) the number of MCR species and the percentage of species whose removal leads to a change in temperature values; and (b) the number of taxa and number of individuals for the MCR subset and full dataset.	199
Figure 6.32. BugStats EcoFigs for the last 20 ka in 1 000 year slices, using taxa data only, (a) using all taxa, and (b) species identifications only.	202

List of tables

Table 1.1. Typical layout of a Quaternary data cross-tabulation, referred to as <i>countsheets</i> in BugsCEP and this thesis.	3
Table 2.1. Summary of BugsCEP program features.....	14
Table 2.2. The de-normalized central INDEX table in BugsCEP. The headers are field names, and the rows are records.....	19
Table 2.3. Hypothetical section of a more normalized version of the BugsCEP index table. UID = Unique Identifier, the key field for this table. The AUTHORITY field could theoretically be normalized further to remove empty cells by splitting it off into a separate table.....	19
Table 3.1. Example of size data stored in the measurable attributes table.	35
Table 3.2. Bibliography table fields, with example.....	36
Table 3.3. Two biology records for the water beetle <i>Ilybius vittiger</i> (Gyll.).....	37
Table 3.4. Available sample date range and uncertainty flags.	42
Table 3.5. Summary of BugsCEP data, showing number of records per data area (support tables excluded).	43
Table 3.6. Physical disk space size of BugsCEP database (see text for disclaimer).....	43
Table 3.7. Available countsheet context descriptions and enumeration in BugsCEP.....	45
Table 3.8. International distribution of known Bugs users.....	50
Table 3.9. A summary of the BugsCEP main interface areas. I/O = Input/Output.....	51
Table 3.10. Red Data Book classification systems available in BugsCEP.....	57
Table 3.11. Basic outline of the ecology classification systems used in BugsCEP.....	59
Table 3.12. Outline of standardization options in BugStats.....	74
Table 3.13. Extract of the numerical results output for the thermal reconstruction from the Saint Bees site, using the ‘Closest to 100%’ option. See Table 3.14 for an explanation of column headers.....	81
Table 3.14. Explanation of MCR results terms, see Chapter 5 for more details on the method. All results are in degree Celcius.....	81
Table 3.15. Data areas currently searchable with the Bugs Search Explorer, along with the method of criteria specification.....	84
Table 4.1. List of Bugs EcoCodes, with diagram label and short description. Indicator classes are in <i>italics</i> , and * marked classes represent narrow habitats. See text for further explanation.....	102
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.....	103
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.....	117
Table 5.1. Explanation of MCR results terms, repeated from Chapter 3. See chapters 3 and 6 for worked examples and outputs. The reconstructed temperature range limits are collectively referred to as TValues.....	121
Table 5.2. Species list return from Predictions module when searching for species with TMin and TMax spans equal to or narrower than <i>Diacheila arctica</i> (Gyll.).....	128
Table 5.3. List of species, and their thermal tolerance extremes, which have TMax ranges equal to or narrower than 7-14°C. Species that would not be able to cope with an increase in summer temperature ranges of 2°C are in bold	129
Table 5.4. List of species, and their thermal tolerance extremes, which have TMax ranges equal to or narrower than 9-16°C. Stenothermic species that would not have been able to survive a 2°C lower TMax value are in bold	130
Table 5.5. Warm/Cold reference cells used in RECON. Reproduced from Perry (1986:131).....	134
Table 5.6. Jackknife statistics for sample s50, Saint Bees (Coope & Joachim, 1980), calculated from the results shown in Table 5.7. See Table 5.8 for more details.....	136
Table 5.7. Jackknife process output for sample s50, Saint Bees site, slightly modified as follows: Species whose removal causes a change in the TValues are in <i>italics</i> ; species whose removal leads to a 100 % overlap area are in bold	138
Table 5.8. Statistics calculated by the BugsMCR jackknife routine.....	139
Table 6.1. Sampled landscape units summarized from Koivula <i>et al.</i> (2004).....	143
Table 6.2. Modified Sørensen’s coefficient of similarity (Southwood, 1978), showing slightly closer similarity between the Edge fauna and the Farmland fauna, than the Edge fauna and the Forest fauna.....	147
Table 6.3. MCR thermal reconstruction for Koivula <i>et al.</i> (2004) reconstructed grouped trap data. All groups produced 100 % overlap regions. Present day climate for southern Finland given for comparison. See Chapter 5 for explanations of the reconstruction method and variables.....	148
Table 6.4. Comparison of number of species per group/classification. Note that BugStats allows species to fall into more than one class.....	149

Table 6.5. Comparison of Ponel's ecological requirement categories and Bugs EcoCode equivalents. Bugs EcoCodes that are not directly equivalent to Ponel classes are shown in <i>italics</i> .	151
Table 6.6. Correlation matrix showing modified Sørensen's coefficient (Southwood, 1978) between La Grande Pile samples. Darker shading represents greater similarity between sample pairs, and coefficient values are on a scale from 0 = no similarity, to 1 = total similarity.	161
Table 6.7. Hemavan – processed sample depths and descriptions. Sample depths are in cm.	168
Table 6.8. Beetle species list from Hemavan (continued on next page).	169
Table 6.9. Thermal limits for species in Hemavan bottom sample 145-160cm. See Chapter 5 for explanation of variables.	173
Table 6.10. Beetle species from Lockarp 7B feature 14495.	177
Table 6.11. Beetle species from Lockarp 7B feature 26551.	178
Table 6.12. Summary counts and sums for the Lockarp 7B 'wells'.	181
Table 6.13. GUS modern, vegetation zone field descriptions.	184
Table 6.14. GUS modern species list, zone abundance sums for the four week collection period.	185
Table 6.15. List of sites used in the 20 000 ¹⁴ C year, 1 000 year time slice reconstruction, indicating the time slices for which each site has ¹⁴ C dated samples. Note the absence of time slices 18-17K, 17-16K and 16-15K which produced no data. See BugsCEP for references for all sites.	190
Table 6.16. Summary of samples and species occurrences from 1 000 year time slices for the past 20 000 ¹⁴ C years. An 'occurrence' is a fossil record of a specific taxon in a specific sample, and may either be an abundance or presence value, see Table 6.16 for more details. Note the important difference between the number of taxa (NSpec), and the number of taxa available for MCR (No. MCR Taxa).	192
Table 6.17. Explanation of cell and total values in Table 6.16.	193
Table 6.18. Modified Sørensen (Southwood, 1978) correlation coefficients comparing 1 000 year time slices for the past 20 000 years. Note that the full dataset has been used, and not just the MCR species, and time slices with no data have been omitted.	200
Table 6.19. The effect of standardization (Std) on the relationship between <i>sample habitat sums</i> and <i>the numbers of species in samples</i> (species level identifications only). Values are R ² , the Pearson product moment correlation coefficient, indicating the proportion of the variance in the habitat class values that is attributable to variance in the numbers of species. Values are shaded by magnitude, and a higher value indicates greater correlation.	203
Table 6.20. Amount of variance per habitat group explained by the different temperature values output from BugsMCR. Cell values are R ² , and <u>underlined</u> numbers indicate negative relationships. TMaxDiff and TMinDiff are the TMax and TMin spans respectively (e.g. TMaxHi-TMaxLo). Cells are shaded by R ² value, see the text for explanation of the bold highlighted cells.	205

List of abbreviations

Bugs	– A general reference to any version of the software created in the Bugs project, concerning the long-term development of a database for palaeoentomology.
BugsCEP	– The latest version of the Bugs software, the Bugs Coleopteran Ecology Package.
BugsMCR	– The climate reconstruction component of the BugsCEP software.
BugStats	– The environmental reconstruction and statistics component of the BugsCEP software.
GIS	– Geographical Information Systems.
(G)UI	– (Graphical) User Interface.
MAL	– Environmental Archaeology Lab, Umeå University (Miljöarkeologiska Laboratoriet).
MS	– Microsoft.
(R)DBMS	– (Relational) Database Management System, a software package for developing and/or administrating databases.
SQL	– Structured Query Language – a common language for database data retrieval.
VBA	– Microsoft's Visual Basic for Applications programming language.

Database terminology

Database	– A collection of related data items, and the information describing them.
Table	– A subset of the database, containing records and fields of more closely related data.
Record	– An individual row of related data in a table.
Field	– A column in a table, where each record/field cell contains the value of the field for that record.
Key Field	– or Primary Key – a field that holds unique data that can be used to identify any record in the table.
Index	– Generally a sorted field in a table, allowing records to be arranged according to some defined sort order such as ascending/descending.
Query	– The database name for a request for specific data from the database. SQL is a type of language for writing these requests.
Backend	– The actual data part of the system, stored on the user's machine or a server.
Frontend	– The software used to interface the data (backend), stored on the computers of individual users, but could also be accessed through a web browser.

General terminology

In any multidisciplinary project there is inevitably a potential risk for the mixing of terminology, and this thesis is particularly at risk by drawing from areas of Quaternary science, ecology, archaeology, geography, computer science and software development. I have tried my best to be consistent by using the definitions favoured by Quaternary science and the consumer end of software development. The former is itself an implicitly multidisciplinary field, and thus has evolved a language which is common to the majority of those who work with the study of human interactions with the environment, including archaeologists. By leaning towards the consumer end of software development the intention was to limit the use of technical development and systems analysis terminology, that is to say, use words that the majority of only slightly computer literate readers should be able to understand. On a similar theme, the use of applied statistical terminology could cause confusion due to the duplication of terms, primarily related to sampling in archaeology and Quaternary geology. The following definitions should aid clarity.

The word "site" is used in its archaeological and palaeoecological definition, as the location of a sampling activity. For example: an archaeological excavation; a lake or peat bog where samples have been taken. Note that this is not the definition used by Jongman *et al.* (1995).

The word "sample" is used in its archaeological and palaeoecological definition, as the actual physical unit of analysis within a core or from a site. For example: a five centimetre high, five litre block from a peat bog, which is part of a column of samples (see Hemavan example in Chapter 6); a one centimetre slice from a lake core; the contents of a bowl excavated from a Norse Farm in Greenland. Note that this is not the statistician's definition of a sample, but sometimes overlaps this.

The terms "clustering" has been used as in Jongman *et al.* (1995), to describe the grouping of points, be they species, samples or sites as defined above. "Classification" has been used synonymously to describe the process of assigning species to specific habitat groups, and the habitat groups occupied by a species. The distinction has been made between *habitat groups (or types)* as definitions of a particular environment or biotype from its physical and vegetative properties (e.g. wetland, dung), and *species groups*, or what Eyre & Luff (1990), among others, call '*habitat groups*' – species groupings that have been compiled through statistical analysis. This is a particularly important distinction to bear in mind when reading Chapter 4.

Although several parts of this thesis discuss statistics, the use of terminology from the field has been limited, as it tends to be at a tangent to palaeoecology terminology. As the latter becomes more and more saturated with quantitative methods, however, this is becoming less the case.

Where the software is described, on-screen buttons to be pressed have been placed in square brackets: [Button], whereas keyboard key presses are indicated by enclosing less than/greater than signs: <F1>.

Database field names are referred to in the case that they have in the database structure, i.e. 'taxonomic CODE' refers to the 'CODE' field in the database, whereas 'taxonomic code' refers to a taxonomy coding system independent of the database.

When a species name is mentioned for the first time in a section it is written in full, whereas the genus is abbreviated on subsequent mentions, e.g. *Carabus nemoralis* Müll. and *C. nemoralis* Müll. This convention is broken when its use could lead to confusion, and when a reasonable amount of text has passed between uses. Authorities (the abbreviated name following the species) are always given to avoid taxonomic misunderstandings, but may have been omitted from a few tables where space was limiting.

A number of aspects discussed are relevant to both palaeo and modern studies. Where this requires emphasis, the 'palaeo' prefix has been bracketed, as in '(palaeo)ecology'. Similarly, where both quantitative and semi-quantitative methods are implied, the form '(semi-)quantitative' has been used.

The structure of the thesis

This thesis by no means follows the traditional structure of a Faculty of Arts work. The combination of software development, which entails an amount of developmental and descriptive text, along with methodological development and then application of the techniques has led to a three part structure, distributed through six chapters.

After the introduction (Chapter 1), there follows in Chapter 2 an account of the development of the Bugs Coleopteran Ecology Package (BugsCEP), including a discussion of the realization of project goals, and a brief developmental history. Although this chapter may appear to be of little interest for many readers, it but puts the rest of the work in perspective.

Chapter 3 goes on to describe first the data within BugsCEP, and then the software tools which have been programmed to allow the entry and use of these data. The necessity of various aspects of the data are discussed, and practical instructions given as to the use of the software, with examples where relevant. This chapter also introduces the sub-components of BugsCEP: BugStats and BugsMCR, which are described in more detail in the subsequent two chapters. Chapter 4 describes the BugStats environmental/habitat reconstruction and statistics software component, putting it in the context of previous work and explaining the methods used in detail. Chapter 5 explains the implementation of the Mutual Climatic Range (MCR) method for climate reconstruction in the BugsMCR software component, explaining some of the refinements made and possibilities for future enhancements.

Chapter 6, which precedes the final conclusions, applies the software described in the earlier chapters to practical examples. The data used in this chapter are a combination of work by myself and work published by others. This chapter may be of most interest to the general reader, although reference to earlier chapters is recommended for a more complete understanding of the methods used.

Acknowledgements

I came to Sweden in 1996 seeking adventure and a PhD. Adventures soon followed, but the PhD didn't progress quite as quickly as I originally intended, but now, more than ten years later the thesis is finished. My primary goal has always been the further development of the Bugs software, which may explain why parts of the thesis are somewhat like a manual. In the original PhD proposal, the software development was to be combined with my own palaeoentomological investigations into past environmental and climate changes in Sweden, preferably along two transects in the north of Sweden. It became evident, however, at an unfortunately late date, that the simultaneous development of the software and comprehensive lab-based analyses was infeasible. As a result, there are only three and a bit of my own sites included in this thesis. The addictive nature of software development is perhaps a

course designed to prevent developers from ever having much time to really use their software. There is always ‘just one more function’ that can be added, and a number of the tools found in BugsCEP are the result of this. As curses go, I don’t think it was such a bad one, and I think the release of a cohesive software package is an important part of this work. Besides, I am reasonably pleased with the release version... for now.

BugsCEP and this thesis would, of course, not have been possible without the help of a large number of other people. Paul Buckland, my father, as the co-founder of Bugs along with Jon Sadler back in the 1980s, deserves a special mention. Without his ideas, encouragement and continual addition of data, the software would never have become such a comprehensive package. I am also certain that his reputation has helped me to do many an interesting thing in the name of science – and that includes white water rafting in Greenland. Yuan Zhuo Don, who programmed the first successful MS Windows version of Bugs, may be solely responsible for tuning me in to the power of MS Visual Basic, and its ability to work in tandem with MS Office. Even with effective tools, development is not always easy, and members of the tek-tips.com forum have repeatedly helped me through programming or database troubles over the past years. Good software rarely gets produced without rigorous testing, and Malcolm Greenwood, Fredrik Olsson, Eva Panagiotakopulu and Nicki Whitehouse have been especially forthcoming with comments on various bits of BugsCEP. Eva has long been supportive of the Bugs work and contributed significantly to EgBugs, and Fredrik in particular became an avid tester of the BugStats component. Geoffrey Lemdahl, my unofficial external supervisor in the latter part of the PhD, provided many useful suggestions and the preliminary identifications from Njulla. Frank Köhler was kind enough to provide me with a copy of the German Coleoptera database, which proved invaluable in sorting out a number of taxonomic problems. My father, Paul, and fiancé Tina have helped considerably in making the thesis more understandable, and both worked hard to ‘deswenglify’ it. I am also most grateful to Sheila Hicks, who has repeatedly made the time to discuss various aspects of the thesis, science and life with me during our somewhat random encounters. BugsCEP would, of course, be nothing without its data, and although there is no room to thank well over 1 000 authors here, a short thank you list is on the website www.bugscep.com. I’m sorry if I have missed anyone, but please send me an email and I’ll add you to the website!

Despite many fine words, funding for interdisciplinary research is much harder to find than for single science studies. I have been fortunate enough to have my PhD position funded by The Bank of Sweden Tercentenary Foundation, under the Northern Crossroads project, and this, together with a number of smaller European Science Foundation and Faculty of Arts grants have made it all possible. My parents and Grandfather also financed several trips to the UK, for which I am eternally grateful. Matti Koivula and Philippe Ponel kindly permitted me to use their data in the thesis case studies, and Peter Rosén worked hard to obtain and help process the, as yet unfinished, Njulla cores.

As a PhD student I have had the opportunity to study a number of external courses, and three of these proved particular fruitful. Firstly, ‘Biostatistics’ at EMG, Umeå, after which Tom Korsman suggested the multiple removal jackknife technique. Secondly, ‘Jackknifing and bootstrapping with applications’ at SLU, Umeå, and especially the discussions with Magnus Ekström on the use of jackknifing with MCR. Finally, the ESF HOLIVAR ‘Quantitative climate reconstruction and data-model comparisons’ course at UCL, London, which provided much inspiration. I’m a firm believer in never working within disciplinary boundaries, and the Environmental Archaeology Lab, Umeå, has not only provided an atmosphere conducive to this theme, but also colleagues and friends that have supported me through the years in a land that was supposed to be full of Vikings, but apparently isn’t. They have even helped me to learn a language where the words for ‘lizard’ and ‘cultivate’ are only an umlaut apart. Foremost of these individuals is my supervisor Roger Engelmark, a true multi-disciplinarian, who has given me the freedom to wander a somewhat free-range course towards the conclusion of this thesis, just in time for his retirement. Johans L and O have also provided many a feedback session, even if it took them a bit too long to understand that a computer cannot properly function with less than two monitors. In addition, Carolina, Helena, Jan-Erik, Karin, Nina, Sara, Sofia and Åsa have all been great company and provided many cakes. Although my research has been somewhat confusing to many at the department of Archaeology and Sámi studies, Umeå, the engagement of staff and other PhD students in seminars and discussions has provided a considerable amount of help and inspiration. Also, we the

PhD students have had our troubles, and I'm very grateful for the mutual support that we have managed to maintain as a doktorand group, Birgitta probably having put more effort into keeping things going than most of us.

I've made a number of very good friends in Sweden, a surprisingly large number of whom seem to understand why somebody would want to remain a student for the equivalent of nine years. There are also a few of them who understand Neil Innes' famed words, "I've suffered for my music, now it's your turn", better than most. Just wait till the Hammond arrives. Anders, Brendan, Janne, Johan O, and PJ – you know who you are. Just for old times sake, Claudia, Markus, Patrick, Sofia and a number of others without whom the little town of Umeå would have been far less fun! Last, but never least, I'd like to thank my family (change camera angle), for their constant support during my studies and adventures in the Arctic wastes. Without my mother, Joan, Dad would never have been able to maintain his enthusiasm for beetle work, as a man is not able to live on cheese on toast alone. She has also constantly supported me in my wanderings. Rob (Bro) deserves some kind of thanks for periodically bombarding me with inane web links, and cute pictures of his kids. In the grand tradition of saving the best to last, I'd like to say at least eleventy million thank yous to Tina, who is without a doubt the best sweetheart in the world! :)

“In this great future, you can’t forget your past”

(No woman, no cry)

Bob Marley – musician, Rastafarian, and so it would seem, Quaternary scientist.

“I’ve suffered for my music, now it’s your turn”

(The Idiot Song, live)

Neil Innes – musician, comedian, and unknowing writer of the above one line PhD summary.



Dedicated to Penguins everywhere.

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.

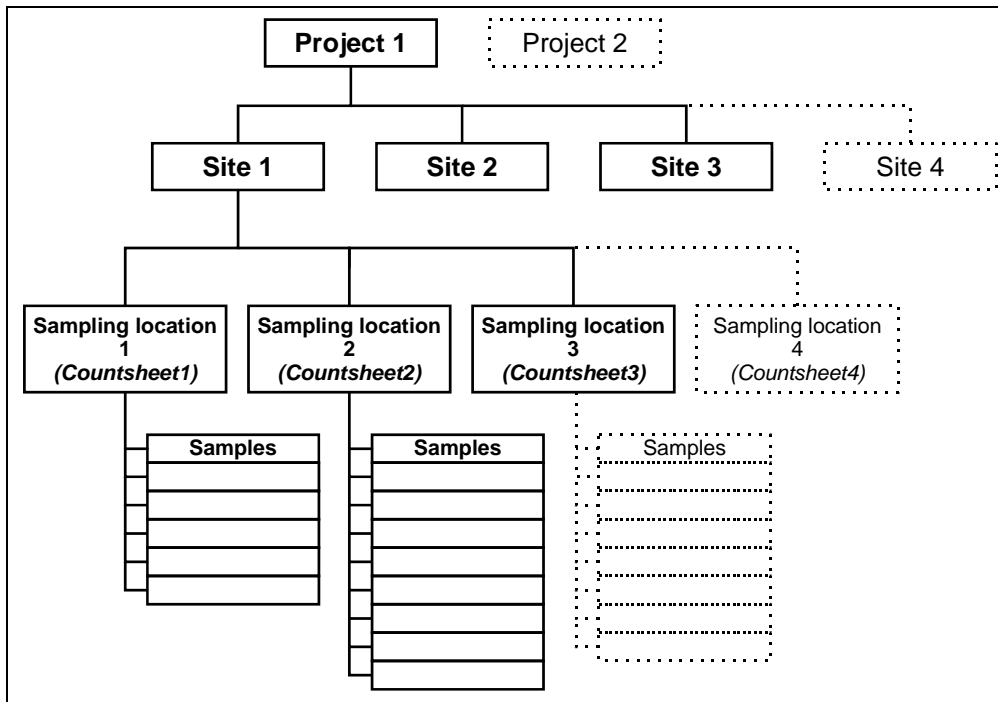


Figure 1.1. Typical Quaternary science sample hierarchy from project to sample level. *Italics* show where the BugsCEP name differs from the common usage. A sample contains abundance data for each species found in it (see Table 1.1).

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.

2 The Development of BugsCEP

This chapter provides a summary of the developmental work undertaken in the creation of the BugsCEP database and its program components. It also provides some basic background information that is a prerequisite for understanding certain more technical aspects of the system. Detailed descriptions of data and interfaces can be found in Chapter 3, and more specific details of the BugStats and BugsMCR components are to be found in the subsequent chapters.

2.1 Introduction

Programming is an undervalued activity in many areas of work outside of those where its application is most obvious – such as modelling, computer science and software development. Although the *data* is what most would regard as the core of any database, it is of limited worth without the software tools to enter, manage, retrieve and query it. In the case of BugsCEP, which is a comprehensive research and teaching software package built around a database, there is a considerable amount of coding, interface construction, query and report design built on top of the actual database. If one ignores the database structure, controls and properties and just examines the surrounding amount of physical (i.e. programmed or written) information it equates to over:

30 840 lines, 135 448 words, or over 892 834 characters
(excluding spaces = 6.6 characters per word on average)

This is the equivalent of over 330 pages of text at 400 words per page. When the numerous rewrites, experiments and adjustments are accounted for, this forms the bulk of the spent time behind this thesis. Word counting the actual data is difficult, but it can reasonably be estimated to over 1 100 000 words. This represents over 2 500 pages of text that has either been converted from earlier versions of the database or entered by one of the current or previous developers. Approximately 60 % of these words is part of the biology and distribution data, a fact which clearly illustrates the importance of modern reference data in the database.

2.2 Database and Software Background

2.2.1 Relational database design

There is considerable variation in the use of the term ‘relational database’. The original, and in some eyes official definition is complex and relies heavily on an understanding of set theory and logic (see Codd, 1970, for the initial definition). A simple layman’s definition is chosen here: a relational database is where similar data are stored in matrices (tables) which are linked to each other by rules (relationships) governing their “...derivability, redundancy, and consistency” (Codd, 1970, out of context). This terminology contrasts with some more formal definitions, which will not be discussed here in order to avoid confusionⁱ. The software used to construct and maintain a relational database is referred to as a ‘relational database management system’ (RDBMS, or RDMS). There is some argument as to which software packages may be considered as RDBMS’s, but for the sake of simplicity MS Access, the database management software used to create the BugsCEP database, will be considered as an RDBMS in this thesis.

A relational database is then, a database designed so that there is no ambiguity over access to data items, no duplication or redundancy, and where data items are linked through a logical system of primary keys and indices which also enforce integrity rules on the data. (Redundancy is the repetition of identical data in a table – in other words, the inclusion of more data than is strictly necessary). By

ⁱ Readers are directed towards the Internet for thorough descriptions of relational database concepts.
http://en.wikipedia.org/wiki/Relational_database is probably a good starting point.

following these guidelines a database can be constructed that is robust and secure in terms of read/write/delete access, and where relatively simple query expressions can be used to retrieve any combination of linked (related) data items. The process of reorganising data structures into relational database form is known as *normalisation*ⁱⁱ.

Quaternary palaeoecology and environmental archaeology data are by their nature multi-proxy, consisting of several measurable quantities that can give a greater insight into past climates, environmental change and human impact together than they can independently. They are also chronologically and spatially extensive, and thus form a multi-dimensional data domain which must be normalised into an efficient data structure if it is to be accessed efficiently. Although the domain is complex, it is logically structured, and thus lends itself relatively easily to normalisation. The Coleopteran data form just a small part of the data-space occupied by the wealth of proxy data sources available, and multi-proxy databases such as SEAD (Buckland *et al.*, 2006) are considerably more complex than BugsCEP.

Database tables are linked by ‘one-to-many’ relationships where the data held in the *primary key field* are unique for every record in that table (the ‘one’ side of the relationship). The field with the same name in the table on the ‘many’ side of the relationship may contain several records where this value is identical, and is referred to as the *foreign key*. In the structural example shown in Figure 2.1, taken from the more thoroughly normalised SEAD environmental archaeology database, the use of one-to-many relationships can be seen for a number of fields. In the table *tblTaxaInsectGenera*, which stores the scientific names of insect genera, the beetle Genus ‘Carabus’ could have the unique identifier GenusID=3 (i.e. in the single record where the field ‘GenusID’ has the value ‘3’, the field ‘GenusName’ contains the value ‘Carabus’). The one-to-many relationship with table *tblTaxaMasterInsects* thus allows for several records in the latter where GenusID=3, the remaining fields of which hold data (or references to data) on all the species within the Carabus Genusⁱⁱⁱ.

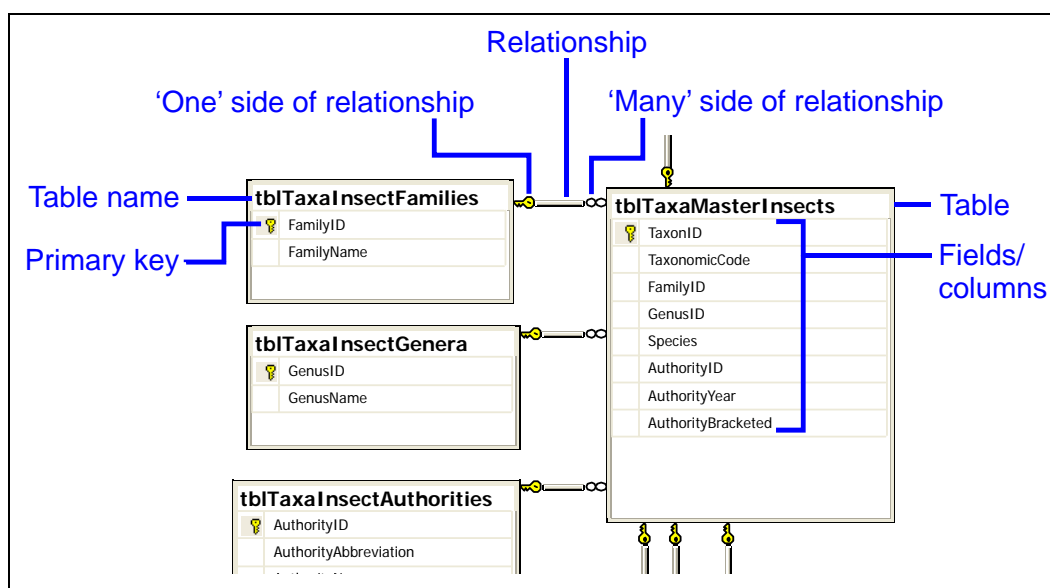


Figure 2.1. Relational database terminology, and the normalized taxonomic index structure of the SEAD database (Buckland *et al.*, 2006).

ⁱⁱ See e.g. Gifford *et al.* (1996), although any practically any relational database design book will do.

ⁱⁱⁱ Note: This structure diagram was created with MS Visual Studio, using the SQL Server Express Edition management console, which can produce more illustrative structure diagrams than MS Access – the RDBMS used to create BugsCEP. The BugsCEP database structure has been recreated in SQL Server to improve presentation, but as yet only runs through MS Access.

Relational integrity enforces rules for the relationship between a primary key field ('one' side of the relationship) and its use as a foreign key in another table ('many' side of the relationship). Once enforced, there can be no data items in the foreign field that are not present in the key field, and it is a useful system for ensuring data validity. In MS Access this function can be left turned off if necessary, thus allowing for the easier import of old data. It can then be activated once the imported data have been verified (and will produce an error if the data violate the relationship criteria), a function which proved very useful during the transference of data from the previous version of Bugs. With referential integrity enforced, changes to any data that are in a primary key field will cascade to all related tables, and thus further ensure data validity.

The relational structure also forms a framework for the construction of database (SQL) queries, which are the primary method of data retrieval, with the logic of the relationships between tables providing pathways for the recovery of any data item. RDBMS packages such as MS Access and MS SQL Server provide intuitive interfaces to assist when creating queries, and write most of the SQL code themselves – a considerable improvement of older database management software. These interfaces, however, are difficult to use without a knowledge of database design, and are replaced in BugsCEP with custom built user interfaces.

2.2.2 The BugsCEP database structure

As described earlier, BugsCEP uses the frontend-backend application structure common to many database applications, as illustrated in Figure 2.2. This setup has many advantages, but perhaps the most important two are:

1. Increased data security – the potential for data loss and file corruption are reduced by keeping the data and program separate. This is particularly useful due to a number of bugs in MS Access, which could potentially cause this to occur.
2. Updatability and ease of distribution – the frontend can be updated without requiring a new copy of the backend. Likewise, updates to the data can be distributed independently of the program files. This means that smaller files can be distributed for updates. It also allows databases to function more easily in a multi-user environment, where a copy of the frontend is installed on each client workstation, but only one copy of the backend database is installed on a central server. In this way all users have access to the latest data^{iv}.

MS Access creates single file databases – that is to say that all the tables, queries, interfaces etc. can be stored in a single file^v. In the BugsCEP system this results in two files, one each for the frontend and backend. The frontend also utilises several small external library files, which fulfil a number of common Windows tasks such as browsing/opening/saving files, and provide access to a number of MS Visual Basic and data management programming language components^{vi}.

^{iv} Note that BugsCEP is not specifically developed to run in a multi-user environment, even though previous versions were. This is a result of the increased complexity and reliance on external library files in the new version, and the fact that there has not been time to put in place the necessary program parts to ensure reliable multi-user access.

^v With the extension '.mdb'

^{vi} Specifically, MS: VBA; Access Object Library; DAO; stdole; ADO; Common dialog library; Excel; Scripting library; Common controls; Hypertext help library.

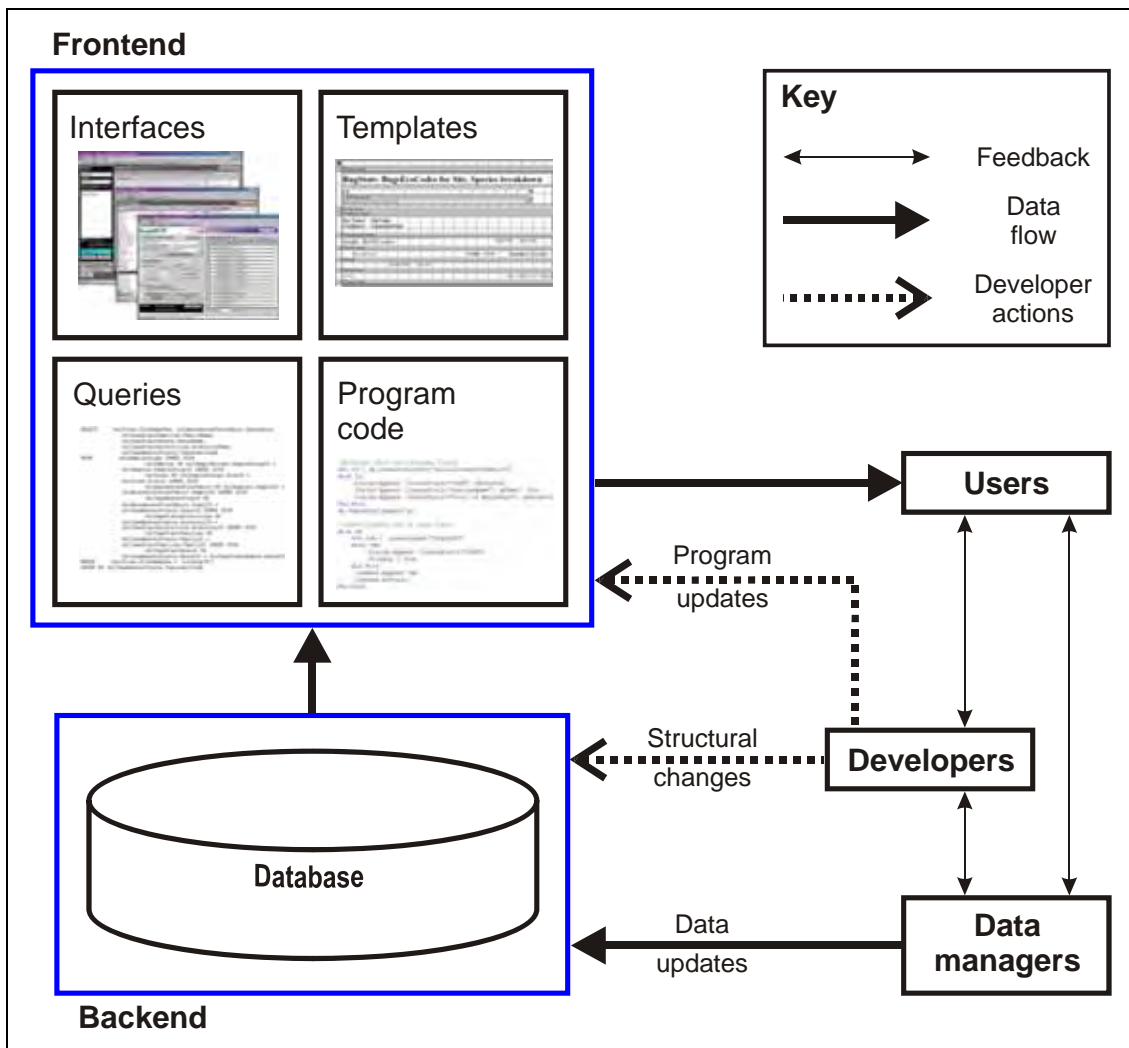


Figure 2.2. BugsCEP program frontend-backend structure, showing user, developer and data manager interaction.

The frontend contains only temporary data tables, along with settings for the current installation, and uses links to retrieve data from the backend. The structure of the backend is a highly, but not fully, normalized relational database as shown in Figure 2.3 (see later in this chapter for a discussion of some of the problems with this structure). The figure shows the relationships between the tables that store the BugsCEP data. An important feature of BugsCEP is its storage of modern ecology and distribution data – rather than just fossil data. This is illustrated on Figure 2.3 by grouping the tables that hold modern or fossil/site based data. Palaeoentomology works by comparing fossil assemblages with the habitats, biology and distribution of modern individuals in order to reconstruct past environments. As a user extracts information for a particular species found in a particular sample, the relationships enforced between the tables ensure that the modern reference data for that species can be obtained directly, and that any changes cascade automatically between the tables.

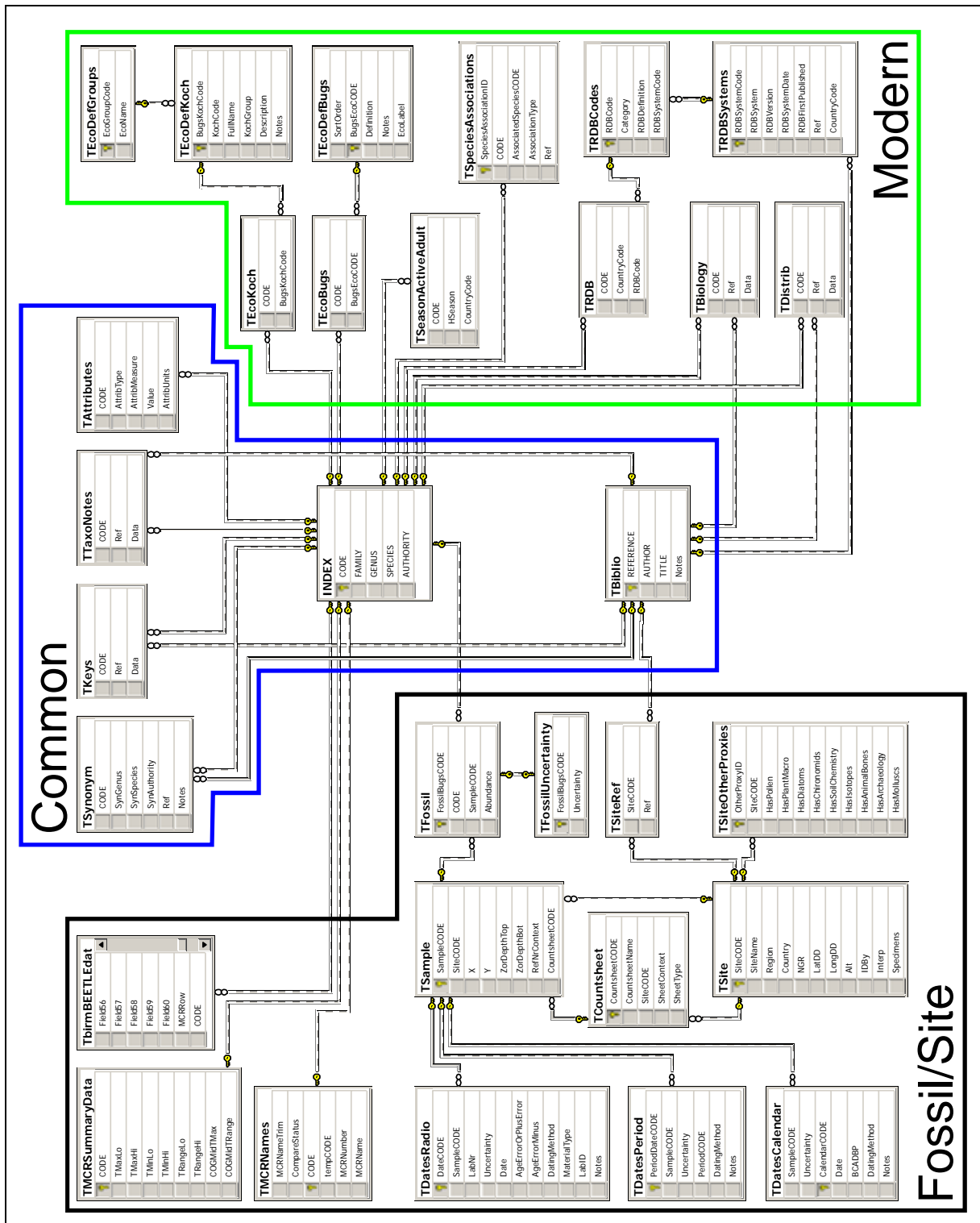


Figure 2.3. BugsCEP backend (database) structure. Boxes represent data tables with lists of their fields enclosed. Lines represent relationships between the tables (see Figure 2.1 for explanation). Note that a number of lookup and reference tables have been omitted to improve clarity. The three table groups, or data areas, Common, Modern and Fossil/Site are explained in Chapter 3.

2.2.3 Overview of BugsCEP software features

BugsCEP contains a large number of features, which are summarized in Table 2.1, and described in detail in the subsequent sections and chapters.

Table 2.1. Summary of BugsCEP program features.

<p>Species data retrieval</p> <p><i>The main interface allows you to browse by species and read or extract:</i></p> <ul style="list-style-type: none"> - published ecological information with references - published distribution data with references - synonyms - RDB - Red Data Book status (only UK so far, but facilities for more) - known (Quaternary) fossil record, sites and dates (if present) - coded ecological summary (Koch (1989-92) and internal Bugs EcoCodes) - taxonomic notes and limited size attributes <p>Search by habitat (and more)</p> <p><i>Lists of species that match specified criteria can be obtained (ecology codes, RDB and biology and distribution text). These lists may be exported with a variety of information - including ecology and distribution, EcoCodes and references. BugsCEP can produce summarised lists of sites which contain the selected species.</i></p> <p>Site and collection data storage</p> <p><i>Users can store their collection/sample data as any number of countsheets per site. Site summary information, including latitude and longitude can be stored, and output in reports. BugsCEP can create a number of reports that summarise/list the ecological and distribution data for all species found at a site, with references. Species lists and countsheets can also be imported from MS Excel files. Lists are automatically sorted into taxonomic order.</i></p> <p>MCR Climate reconstruction</p> <p><i>Mean summer and winter temperatures can be reconstructed from species lists, sample by sample, using the MCR method (Atkinson et al., 1978). BugsCEP will produce simple thermal diagrams in MS Excel, along with exporting the raw data and sample thermal envelopes (climate space maps). It also has the facility to show which species could theoretically survive in any given temperature range, effectively allowing users to predict changes in species distributions with climate change.</i></p> <p>Environmental reconstruction</p> <p><i>Habitat summary diagrams can be created, showing the changing ecological implications of species between samples. A variety of statistical treatments are available – including standardization, abundance weighting and $\ln(n+1)$ transformation. Summary reports can be created to help in the analysis of these diagrams sample by sample. Correlation coefficients can be calculated to assess the (dis)similarity between all samples at a site.</i></p> <p>Reporting and exporting</p> <p><i>A variety of time saving reports can be created and exported, including:</i></p> <ul style="list-style-type: none"> - summary data for any taxon - all information for species found at a site, with references - abbreviated forms of the above - summary information for all sites which include specific species - sample by sample, species by species breakdown of ecological implications of species found at a site - MCR climate reconstructions - Bugs EcoCode environmental reconstructions - exported countsheet in MS Excel format <p>Bibliography</p> <p><i>A comprehensive bibliography of over 3 300 papers on beetle biology, distribution and fossil records is included in BugsCEP.</i></p>

2.3 Developmental Strategy

Database design projects have a greater chance of success if a design strategy is arrived at before the project is undertaken. The strategy may include goals with different levels of priority, but should always encompass an analysis of the system to be replaced, and the goals to be achieved. The goals which orientated the development of BugsCEP are to a large part synonymous with the aims of this

thesis, and are summarised below in order of priority, with comments on the implementation and realisation of these aims.

2.3.1 Primary developmental aims

1. Develop a new, relational database, version of the Bugs software.
2. Develop a system for the graphical comparison and interpretation of fossil/modern insect faunas (with the working name: BugStats).
3. Implement and improve the MCR climate reconstruction method in Bugs, i.e. port to Windows and improve the MS-DOS software, and make MCR generally more accessible to all (with the working name: BugsMCR).

Implementation of Primary Aims

1. A completely new database structure, as described above, and new interfacing software were created to form BugsCEP. To avoid the inevitable legacy problems common when trying to adapt existing software to new purposes, no program code was retained from the old version. The software components are described Chapter 3 of this thesis.
2. & 3. BugStats and BugsMCR were developed as individual programs in order to achieve a more rapid release as test versions, and enable the use of the encompassed methods in teaching. Once fully functional, they were then integrated into the main BugsCEP program, and phased out of use as standalone programs. Such a strategy was greatly eased by the use of object orientated and modular programming, and only a minor amount of revision was needed in order to get the components running together. Redundant program code, that is to say code that performed essentially the same task but was found in more than one program component, was rewritten as common subroutines/functions with the appropriate structure to enable it to be called from the original locations. This also allowed for easier debugging, despite the inevitable increase in complexity of the subroutines. The user would notice no change, but future developers would be extremely grateful for having to only modify the program in one location rather than several. BugStats and BugsMCR are outlined in Chapter 3, and described in detail in chapters 4 and 5 respectively.

2.3.2 Secondary aims

4. Test the system on own data.
5. Test the system on published modern and fossil data.
6. Formally publish and distribute the BugsCEP software, and present it in a medium which allows for extensive discussion.

Implementation of Secondary Aims

4. The fossil Coleoptera remains from two sites were examined: a peat sequence from northern Sweden (Hemavan), two archaeological sequences from southern Sweden (Lockarp). In addition, a small modern fauna from Greenland was re-examined (GUS), and an attempt was made to assess the last 20 000 ¹⁴C years of climate change, as reflected in the faunas from ¹⁴C dated samples stored in BugsCEP. Preliminary data were also examined from lake Njulla in northern Sweden. Two further Swedish sites were examined, but not included in this thesis due to either lack of sufficient fossils or time (Bymyran and 'David's Bog'). The case studies are presented in Chapter 6.
5. Two published datasets were chosen to provide a test of the tools contained in BugsCEP for aiding the interpretation of modern and fossil beetle assemblages, the limited amount of time available restricting the number of datasets that could reasonably be examined. The reanalysis of a recent Finnish study of Carabidae across forest-farmland transects, and a study of a 140 000 year long peat sequence from eastern France, are presented in Chapter 6 section of this thesis.

6. The <http://www.bugscep.com/> website provides an effective medium for the distribution and passive marketing of the database. Additional active marketing, in the form of posts on newsgroups and emails to established contacts announcing the release have been undertaken to a limited extent. The publication of this thesis allows for a more comprehensive discussion of the data and functionality of BugsCEP than individual articles or a manual would allow. With hind sight, an active marketing initiative is necessary to ensure that people are aware of the software and its possibilities, and attain the widest possible user base.

2.3.3 Development platform

BugsCEP was developed using Microsoft Office 2000 Developers Edition in the MS Windows 2000 and XP operating systems. The majority of the work was undertaken in MS Access, which provided easy access to a programming language and database motor, a combination which was not easily available when the project was initiated. The programming is entirely in Visual Basic for Applications (VBA), which is common to all MS Office software, and allows for easy communication between the applications. This enabled, for example, MS Excel to be run from within the BugsCEP code when spreadsheet functions were required, rather than having to simulate them in the database environment. The developmental potential of VBA is considerable, and it has now even been chosen as the programming language within Corel's CorelDraw and ESRI's ArcGIS. This opens up new potential routes for the future development of Bugs in the areas of diagram creation, map production and Geographical Information Systems (GIS).

2.4 Database Structural Changes

2.4.1 Bugs database structure and contents

The structure of BugsCEP is a great improvement in terms of storage efficiency and manageability over the preceding Bugs2000 structure (Figure 2.4, and see also section 2.8.3 for version history). It is these improvements that have allowed for a more flexible user interface, and more advanced searching, data-linking and querying.

The new structure was arrived at by first examining the existing data, and applying the first three normal forms (of normalization) as described by Gifford (1996) and others. This design was then expanded to include everything else that could be reasonably included. The aim was to create a structure capable of holding all the present data, plus any new that were considered important, in a logical structure that minimized repeated information and allowed for the rigorous enforcement of validation and relationship rules. This is standard database development procedure, and can be read about in any number of manuals (e.g. Gifford, 1996) and so will not be described in more detail here. Readers interested in learning more in these areas should turn their attention to Internet searches, or the websites of any major database software developer^{vii}. The final BugsCEP structure (Figure 2.3, Figure 2.5) consists of 43 main tables which contain the data, and 30 temporary tables which hold calculation results and settings. When compared to the structure of Bugs2000 (Figure 2.4), which had 19 data tables and 15 temporary tables, the increase in complexity is unambiguous.

^{vii} E.g. <http://www.postgresql.org/>, <http://msdn2.microsoft.com/en-us/sql/>, <http://www.mysql.com/>, and others.

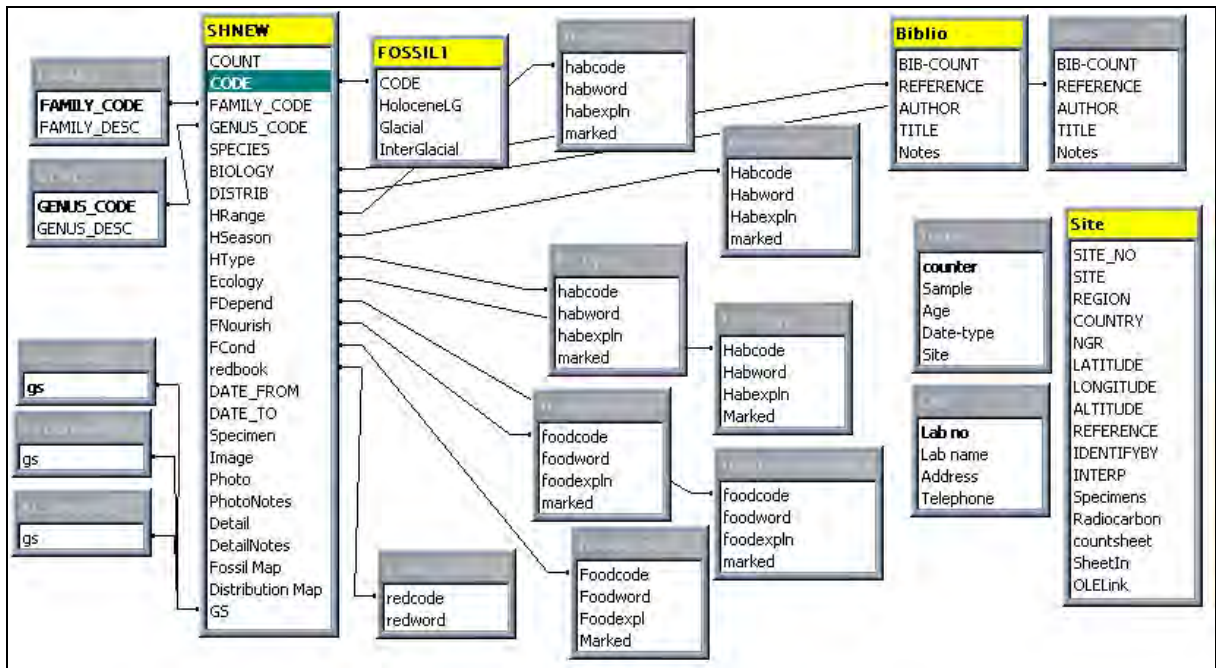


Figure 2.4. Bugs2000 database structure. Note that the relationships between the Biblio and SHNEW are symbolic, and represent programmed links rather than enforced relationships.

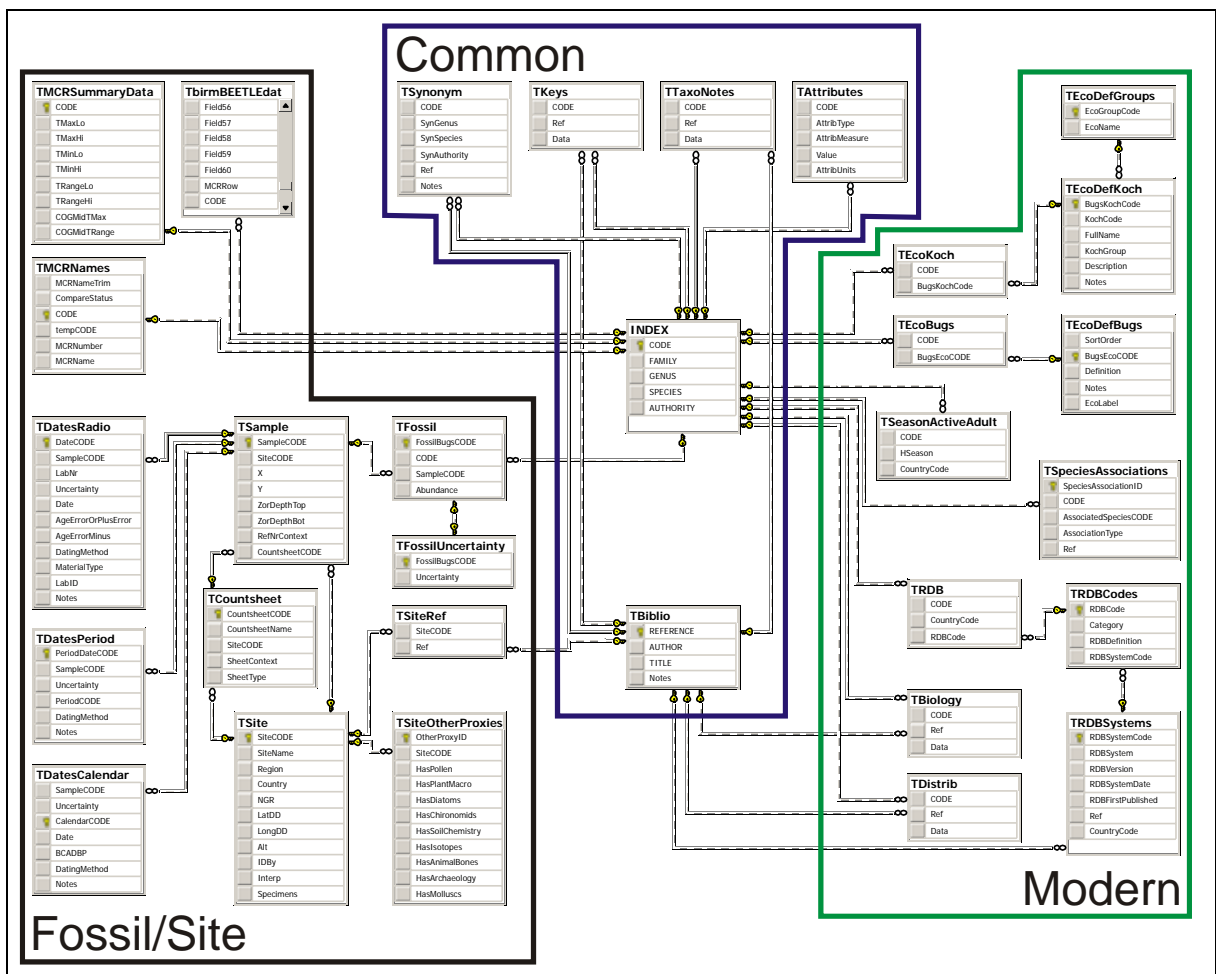


Figure 2.5. BugsCEP database structure, repeated from Figure 2.3 to allow for easier comparison with the structure of the previous version, Bugs2000 (Figure 2.4).

It was decided to drop a number of elements that were provided for in Bugs2000, in order to be able to complete the project within the scope of a PhD. Images and scanned maps were omitted, the former to be re-integrated in a later version using a more optimized external database form. Scanned distribution maps were deemed a poor, although often useful, substitute for a real-time mapping engine or GIS. Although Bugs2000 stored site coordinates there was no system built in for displaying their location on maps. The same is currently true for BugsCEP, although the author has experimented with a simple map displaying interface using VBA, and all the maps in this thesis are produced in ESRI's ArcMAP after exporting the necessary data from BugsCEP. Modern distribution maps, on the other hand, are a far more complex issue, theoretically consisting of thousands of points or areas for a single taxon just for the present day records. Recent developments in online mapping systems (e.g. GEON, Google Maps^{viii}), a number of which allow the free linking of databases, may provide an efficient solution to these problems in subsequent versions of Bugs.

2.4.2 Conversion of Bugs2000 data

The conversion of data from the existing, poorly normalized, database of Bugs2000, into the almost fully normalized BugsCEP database was an enormous and time consuming task. Every item of data from the original version had to be parsed^{ix}, validated, and then manually checked after import. This was particularly problematic for the fossil data, which was stored entirely in notes fields in Bugs2000. After numerous attempts to extract the data, which consisted of an estimated 23 000 fossil records, themselves consisting of site name, reference, date, and occasionally other metadata, this was aborted. It was decided to recreate the data from the bottom up, converting the existing MS Excel countsheet files into the new BugsCEP format, and then importing them into new sites in the new system, parsing them into normalized data on the way. Dates were then manually added in a new Date Manager interface. A positive by-product of this endeavour was the development of the advanced file import/convert system that now forms part of the new interface. In addition, the laborious conversion attempt gave an ideal opportunity to assess the nature of the data, and devise a more optimized date handling system to cope with the wide variety of dating forms and accuracies used in Quaternary science and archaeology.

In Bugs2000, biology and distribution text abstracts were in the form of large memo field data items, where references were enclosed in curly brackets, e.g. {Buckland 2000}, and followed by the text abstracted from that source. Several of these source-data combinations would exist in each record for a taxon. A routine was programmed to extract the individual reference-data items into separate records in the new structure, with a number of validation checks to cope with missing brackets, typos and other inconsistencies. This process led to approximately 19 300 biology records and 16 500 distribution records in the new structure, derived from the original 5 500 records in Bugs2000. The new tables can be rapidly sorted by author as well as taxon, an action that would have been impossible with the old version. The new structure also allows for the references to be linked directly to a master bibliography, from which changes can be universally applied, whereas references had to be programmatically extracted from the memo fields in the old version – a slow and cumbersome process.

2.4.3 Database structural compromises

The BugsCEP database file includes a single de-normalized master *INDEX* table, which includes repetition of family and genera names, as shown in Table 2.2. This is, technically, an inefficient way of storing information, in that the repeated text is redundant information that takes storage space, slows searches, and prevents the enforcement of referential integrity on that field. The more structurally efficient way of storing this information would be in about four tables, one for the FAMILY, one for the GENUS, one for the AUTHORITY, and an index table for the taxonomic CODE and SPECIES with foreign keys relating the other tables (as in Figure 2.1). Each of these tables

^{viii} GEON: <http://www.geongrid.org/>; Google Maps: <http://maps.google.com/maps>

^{ix} I.e. fields containing compounded data had to be chopped up into individual data items.

would have a unique identifier field (UID) that would allow single records for each unique data item to be linked through relationships. The normalized index might look something like Table 2.3.

Table 2.2. The de-normalized central INDEX table in BugsCEP. The headers are field names, and the rows are records.

CODE	FAMILY	GENUS	SPECIES	AUTHORITY
01.0010001	CARABIDAE	Carabidae	indet.	
01.0010020	CARABIDAE	Cicindela	sylvatica	L.
01.0010050	CARABIDAE	Cicindela	hybrida	L.
01.0010060	CARABIDAE	Cicindela	maritima	Dej.
01.0010070	CARABIDAE	Cicindela	campestris	L.
01.0010080	CARABIDAE	Cylindera	germanica	L.
01.0010122	CARABIDAE	Cicindela	sp.	
01.0010125	CARABIDAE	Cicindela	spp.	
01.0020010	CARABIDAE	Calosoma	inquisitor	(L.)
01.0020020	CARABIDAE	Calosoma	sycophanta	(L.)
01.0020050	CARABIDAE	Calosoma	reticulatum	(F.)
01.0020082	CARABIDAE	Calosoma	sp.	
01.0020085	CARABIDAE	Calosoma	spp.	

Table 2.3. Hypothetical section of a more normalized version of the BugsCEP index table. UID = Unique Identifier, the key field for this table. The AUTHORITY field could theoretically be normalized further to remove empty cells by splitting it off into a separate table.

UID	CODE	FAMILYID	GENUSID	SPECIES	AUTHORITYID
1	01.0010001	1	1	indet.	
2	01.0010020	1	2	sylvatica	1
3	01.0010050	1	2	hybrida	1
4	01.0010060	1	2	maritima	2
5	01.0010070	1	2	campestris	1
6	01.0010080	1	3	germanica	1
7	01.0010122	1	2	sp.	
8	01.0010125	1	2	spp.	
9	01.0020010	1	4	inquisitor	3
10	01.0020020	1	4	sycophanta	3
11	01.0020050	1	4	reticulatum	4
12	01.0020082	1	4	sp.	
13	01.0020085	1	4	spp.	

The reason for not normalizing the index, and in fact de-normalizing it from the previous version, is simply for the sake of the convenience of having the full species name available in one table. It simplifies query creation, and VBA coding by having the data available in this way. An alternative would be to normalize the table, and then use a saved query to create a simulacrum of the desired index, which could then be referred to in queries and VBA. In many RDBMS this would be the ideal solution, but unfortunately MS Access has a bloating problem with query use (i.e. the database file grows every time a query is run), and so this alternative has been avoided. Future versions of BugsCEP will use an alternative RDBMS, such as SQL-Server, MySQL or PostgreSQL, which do not suffer from this problem.

2.4.4 Problems and potential problems with the BugsCEP structure

Most of the problems described in this section can be attributed to the inexperience of the developers, and are most likely symptoms of learning by doing, rather than a formal education in database development. They have led to a number of complications in data manipulation or retrieval, all of which have had to be solved by small bits of extra program code.

Use of the *double data type* for taxonomic CODE and calculations

The *double data type* stores floating point numbers, and as such approximates decimal fractions with the closest binary equivalent (computers ‘think’ in binary). This means that some numbers simply cannot be represented by the *double data type*, and while this may generally not be a problem for the taxonomic code itself, it does cause problems where mathematics, and the *double data type*, are used to validate entered codes. Take the following case:

```
Cdbl(Int(CDec(87.0290060) * 10000000) / 10000000) < 87.0290060 = false
```

but

```
Cdbl(Int(CDec(87.0290070) * 10000000) / 10000000) < 87.0290070 = true
```

Clearly *both* formulae should evaluate to false, but they do not as the Cdbl command converts the left part of the comparison to the *double data type* before comparing with the right part.

The following VBA code is used to work around the problem in the validation of new taxonomic codes:

```
Val(Cdbl(Int(CDec(NewCODE) * 10000000) / 10000000)) < Cdbl(NewCODE)
```

Where ‘NewCODE’ is the code entered by the user which is passed to the validation routine as a *double data type* variable.

See Microsoft Knowledge Base (MSKB) article 242933 for more details of this type of problem^x.

Use of taxonomic CODE as key fields

A standard rule of database development is that business data and structural data should not be mixed. Business data is that which has meaning, whereas structural data is that which primarily provides the anchors for the relationships between tables and allows for the efficient use of lookup tables. Taxonomic CODE is clearly business data, and its use as a primary key in the *INDEX* table is clearly a structural function (see Figure 2.3).

Whilst this normally would not cause any operational problems, and in fact makes debugging easier in some respects (as the taxonomic CODE is visible in the majority of tables), it does make the implementation of major changes to the nature of taxonomic codes awkward. For example, the choice of data type used for the taxonomic code field was unwise, as explained above, and it should really be changed. Due to the structural nature of the CODE field, however, this would involve opening the design view of each and every table where the field is present and changing the data type. It is not, however, possible to change the data type of a field that is part of a relationship, so each and every relationship would first have to be deleted. These would then have to be recreated subsequently. These tasks in themselves would only take a number of hours, but then the program code, which has routines created specifically for handling taxonomic codes in the double data type, would have to be checked and adjusted to cope with the new data type. This task could take days, and would then have to be followed up with a comprehensive testing of every part of the program. Changes of such a

^x <http://support.microsoft.com/kb/242933/en-us>

comprehensive nature also usually lead to unforeseen consequences, and would undoubtedly need the addition of new error handling routines. Had an alternative unique identifier been used as the key field instead (cf. the TaxonID field in the *tblTaxaMasterInsects* table, Figure 2.1), then the data type of the separate taxonomic CODE field (cf. the TaxonomicCode field in the *tblTaxaMasterInsects* table, Figure 2.1) could have been changed with considerably less effort. This structure would also allow for the easier implementation of multiple taxonomic systems.

Use of a decimal separator in taxonomic codes

The European Taxonomic Code uses a decimal separator to divide between family and genus numerical equivalents (see section 3.1.1.1). Numbers are, logically, best treated as numbers in databases and program code, and so the taxonomic code is stored and manipulated as such^{xi}. The English language uses a decimal point (‘.’) as a separator, whereas a number of other languages, including Swedish, use a comma (‘,’). Structured Query Language (SQL) can *only* handle a point as a decimal separator^{xii}, and will produce an error if a comma is used. On non-English systems, taxonomic codes are retrieved with a comma instead of a decimal point, and are thus passed to SQL statements with the former, causing a syntax error. The following code therefore is used to replace the comma with a decimal point in codes before they are inserted into SQL statements:

```
CommaPos = InStr(1, CODEAsString, ",")
If CommaPos > 0 Then
    CODEAsString = Left(CODEAsString, CommaPos - 1) & "." &
        _Right(CODEAsString, Len(CODEAsString) - CommaPos)
End If
```

A related problem is that SQL cannot naturally cope with the use of apostrophe’s (‘’) in string values. These must be doubled up in order to not cause an error on processing the SQL.

The following function is used in BugsCEP to cope with any value which includes apostrophes:

```
Public Function SQLEncode(sqlValue As String) As String
    SQLEncode = CStr(Replace(sqlValue, "'", "''))
End Function
```

It returns a version of the passed string ‘sqlValue’ which can be inserted into SQL. For example:

```
SQLEncode("test's") = test''s
```

Use of a limited number of prefix zeros in unique identifiers

Unique identifiers are automatically generated in BugsCEP for all new data items according to a template consisting of a four letter prefix, which designates the type of data, and six numbers, which are increased in order of creation. The six digit number allows a maximum total of 1 000 000 data items in a particular type. This limit seemed more than sufficient at the time of creation, and was in fact one digit, or a factor of ten more than the calculated need. With the current c. 88 953 fossil record data items this limit should theoretically be adequate for at least the near future, considering the small size of the palaeoentomological community. Unfortunately, deleted identifiers are not reused unless they were the last ones created, so due to the revision of lists and a number of experiments the highest fossil record identifier is at the time of writing ‘FOSS098663’. This suggests that the number part of the actual identifier codes will usually be about 10 % higher than the physical number of records, and consequently that the probable limit to the number of records will be about 909 000.

^{xi} Note that other unique identifiers in BugsCEP are either stored as alphanumerics or integer numbers to avoid the problems described here, among others. Of course, these are not without their problems as well.

^{xii} At least in the MS Access implementation.

In the event of upsizing of the database to a more powerful management system, such as SQL-Server or MySQL for example, and assuming a future with prolific palaeoentomology work, it would probably be a good idea to revise this system to avoid limits being reached. There are wide ranging programmatic implications of changing the identifier code system, in terms of code generation and validation, which would take time to resolve, but none are too complex to be a critical hinder. The use of simple, integer identifiers would reduce the potential for future problems.

2.5 Program Development

2.5.1 Summary of deficiencies in the previous version (Bugs2000)

Bugs2000 had a significant amount of room for improvement. Below are listed some of the main limits of the system, from the point of view of the user, although some are essentially structural in nature.

- Synonyms not searchable
- Only one countsheet per site
- Only one reference per site
- Limited number of search terms enterable (three biology + three distribution + three Red Data Book)
- No climate reconstruction (MCR)
- No built in system for quantitatively summarizing faunas, or calculate other sample based analyses
- Fossil record is memo/notes field based, and not searchable
- No facility for storing samples, and associated metadata
- No facility for handling dates linked to samples
- No ability to import imperfect data – i.e. data must be in exactly the right format, and only include the taxa names found in Bugs2000
- Inflexible Red Data Book (RDB) system that only allows UK rarity data to be entered
- Koch EcoCodes interface cumbersome, and based around bad structure
- Taxonomic notes mixed with biology data
- Limited reporting facilities

A number of these limitations have been identified by user feedback, and all of the above have been resolved in BugsCEP.

2.5.2 Main improvements in the BugsCEP system

What follows is a brief annotated list of the main improvements of the BugsCEP program interface and functions over previous versions. Chapter 3 includes detailed descriptions and screenshots of the actual interface parts.

The Main BugsCEP screen

Effective use of tabs to reduce clutter, whilst increasing the amount of information easily available for a species. Synonyms and taxonomic notes are now separated from the main biology data for ease of access. Alternative navigational facilities, including a searchable synonym browser and code browser are provided.

Users with administrator rights can easily switch to ‘Edit Mode’ and make changes or additions using the same tabs, but with editor user friendly functions enabled. A password is required for administrator access, and is available from the author on request. This system reduces the risk of accidental data loss by those not familiar with the program.

Site Management interface

Previous versions had no way of getting a rapid overview of all sites stored in the database, the Site Manager now provides this. It displays summary data for each site on selection from a scrollable list, along with buttons to quickly access the site’s description (metadata), countsheets, references and dating evidence.

Multi-level nesting of data

A site may now hold an unlimited number of countsheets, enabling complex sites to be sub-divided into manageable units, or multiple years of pitfall trapping expeditions to be stored logically under the same site name. A site may now have any number of references attached to it, which is necessary when sites are re-examined, or analysed over a long period of time and resulting in several publications. Unfortunately references cannot be attached to a countsheet, but the latter can be named to reflect their source.

Any number of dates may be ascribed to samples, which themselves are always linked to a particular countsheet. This allows for re-dating information to be stored, as well as a mix of radiometric, calendar (or archaeological) and general period dates to be ascribed to the same sample.

Search system

A new, multi-stage search system has been developed, that allows searching on the majority of data fields, including both by keyword and item selection. Search stages are logged, and the latest step can be undone. Search logs are exported with reports, and the reporting facilities have been extended. Search results can also be exported as a Bugs Countsheet, for re-import into a site. Searches are saved between sessions, so that they can be continued even after restarting the program.

A powerful new function to report the sites where search result species occur has been added.

Internal Site/Countsheets Management

Although Bugs2000 could be used to create species lists, these had to be opened in MS Excel to add samples and abundance data, which was a troublesome and error inviting task. BugsCEP allows for all countsheet management tasks to be performed within the program, including the creation of species lists, sample lists, specification of metadata, and addition of abundance data. It has a number of built in validation routines to prevent the missing of taxa or samples, or the entry of illegal abundance data. This bypasses many problems caused by the free range given to users when entering countsheets data in MS Excel. In the old version, it was difficult to replace a taxon, and one risked losing abundance data if not very careful. BugsCEP provides a button for this.

The new system has been enabled by the normalization of the fossil record system, which now includes a record for each and every presence of a taxon in a sample.

MCR – Climate reconstruction

The MCR (Mutual Climatic Range) thermal reconstruction method of Atkinson *et al.* (1986) has been implemented in BugsCEP, and can be run on any site that is stored in the database. In addition, output of thermal envelopes and samples species lists is provided, as well as a temperature based species

presence prediction system. The latter can be used to help answer questions on the potential effects of climate change on the insect fauna.

More advanced functions, including jackknifing, to calculate the internal reliability of a reconstruction, and relative cold/warm component identification are under development.

BugStats – Environmental reconstruction

Using the new Bugs EcoCode habitat classifications, users can summarize the habitats represented in samples numerically and graphically. A total of 16 possible calculation option combinations are available, including exclusion of generic level identifications, log transformation, abundance weighting and standardization by sum of represented environments. This is potentially a very powerful tool in palaeoenvironmental reconstruction, and provides flexible output in MS Excel format. A report can be produced listing the habitats represented by the species at a site on a sample by sample basis, with abundance values.

In addition, a correlation coefficient calculation system was created, which can output an MS Excel file containing a matrix showing the measure of similarity/difference between each and every sample at a site. Currently only the Bray-Curtis modified Sørensen's coefficient of similarity (Southwood, 1978) is activated, but more coefficients will be added at a later date.

Advanced file importer

The file import system accessed through the countsheet manager includes automatic synonym replacement, including a user customizable record of manual replacements. It also includes an easy to use interface to allow for the resolving of name differences due to typing and spelling mistakes. Converted files are imported directly as countsheets, removing the need for repeated export and import. The files can be exported in the BugsCEP format at any time, once imported.

Multinational and international Red Data Book (RDB) rarity recording system

Records of species rarity (see section 3.1.2.3) are no longer limited to the UK RDB, data can be entered for any number of countries, and for international records. This makes BugsCEP potentially much more useful for studies on biodiversity changes and for monitoring endangered species.

Enhanced reports, and more of them

The standard reports from Bugs2000 have been dramatically improved, now including full references and site summary data, as well as the biology and distribution data for all species found at a site. A number of additional reports have been added, including more flexible export to either MS Word (as an RTF file) or MS Excel. In particular, users can select from a number of variants of a standard report with different levels of information. This facility is available for search results and single species, and will be implemented and expanded elsewhere in the program.

Bibliographies for sites or species can now be exported in report form to either MS Word or Excel.

2.6 Testing

2.6.1 Developer testing

Until 2005, Bugs2000 and BugsCEP, and all components, were tested on MS Windows 98, 2000/NT and XP machines. Windows 98 was dropped at that point due to its increasing scarcity, but as Windows 2000/NT is still considered a stable operating system in many laboratory environments it is

still supported. Due to lack of backward compatibility, a common problem with Microsoft developer environments, the setup files for installation on Windows 2000 must be packaged on a Windows 2000 machine and not a system running Windows XP.

2.6.2 User-based testing

The user test base has primarily been kept small due to two reasons: 1) managing feedback from users during testing of unfinished software consumes a lot of time; 2) there is a tendency for test users to start actually using the unfinished features, which is not advisable. Undergraduate students have been used as in house testers at the Environmental Archaeology Lab, and have provided considerable help in particular with the BugsMCR component.

User testing has been particularly useful in aiding the development of a more intuitive interface, as the developer's technical understanding of the system can often give him a false impression of simplicity in interfaces. That is to say, after having programmed the system for more than five years the developer's understanding of simplicity will be different from anybody else's. In particular, test user feedback has lead to:

- additional explanation of the frontend-backend linking procedure on first run
- help files being included in this release (they were planned to come later)
- the drop down site choosers in BugStats and BugsMCR
- fine tuning of BugStats and BugsMCR output files and reports
- numerous layout aspects, including the use of numbered stages in complex task sequences

2.7 Presentations and Publications

Although there is limited scope for the production of individual papers within a monograph orientated PhD, it has been possible to produce a number of publications. This has been deemed necessary, in that papers are the accepted scientific discussion medium amongst the older part of the scientific community^{xiii}, and that software essentially needs supporting publications to be noticed. Conference presentations have been targeted over journals, in that these presentation and discussion fora provide a greater level, and often more immediate, feedback, and allow demonstration and distribution of the software to interested parties.

2.7.1 Publications about or with direct use of Bugs

Buckland, P.I. & Buckland, P.C. (2002). "How can a database full of Bugs help reconstruct the climate?"

Provides an overview of the BugsCEP development strategy, with an emphasis on the MCR and related data acquisition aspects. Provides a blueprint for a GIS based system for improving the MCR dataset.

Buckland, P.C. [Assisted by Buckland, P.I. & Hughes, D.] (2005). "Palaeoecological evidence for the Vera hypothesis?"

A practical application of the BugStats environmental reconstruction component of BugsCEP, showing a variety of sites useful in elucidating the Vera hypothesis (Vera, 2000) with respect to the British Isles. Summary site lists were produced for this report using custom build queries in the BugsCEP database, and time slice maps then created using ESRI's ArcView.

^{xiii} This is almost certainly less true for the Google-Wiki generation.

Buckland, P.I., Buckland, P.C., Panagiotakopulu, E. & Sadler, J.P. (in press). "A Database for Egyptian Entomology"

An overview of the Egyptian spin-off of Bugs2000 (see section 2.8.4.2), explaining its structure and use.

2.7.2 Unpublished presentations

2002 (1 May) "The Royal Entomological Society. Entomological Computing And Technology Special Interest Group – Recent Advances In Entomological Computing and Technology" London, UK.

Presentation: Buckland, P.I. & Buckland, P.C. "BUGS in the system: a database of Quaternary entomology and Holocene habitats (with a footnote on EGBUGS, an Egyptian version.)".

2.7.3 Poster presentations

2001 (19-22 August) "VIII Nordic Conference on the Application of Scientific Methods In Archaeology" Umeå, Sweden.

Buckland, P.I., Buckland, P.C., Engelmark, R.E., Sadler, J. & Atkinson T. "*Climatic reconstruction through beetle proxy temperature data*". (NB. This poster is the same as the PAGES poster below).

2001 (27-31 August) "PAGES-PEPIII -- ESF-HOLIVAR Past Climate Variability Through Europe and Africa An International Conference". Centre des Congrès, Aix-en-Provence, France. Buckland, P.I., Buckland, P.C., Engelmark, R.E., Sadler, J. & Atkinson T. "*Climatic reconstruction through beetle proxy temperature data.*"

2003 (23 June - 4 July) "Quantitative climate reconstruction and data-model comparisons" Environmental Change Research Centre, UCL, UK. Buckland, P.I. "*Climate & Environmental reconstruction from fossil beetles*".

2005 (13-14 November) "Joint HITE-POLLANDCAL Conference: Human impact on terrestrial ecosystems on long to short term scales with an emphasis on pollen calibration and quantitative reconstruction of past land-cover changes". Umeå, Sweden.

Buckland, P.I. & Olsson, F. "*Using insect analysis and ecological coding to reconstruct human impact and environmental changes during the late Holocene*".

2.7.4 A note on other publications using Bugs (2000/CEP)

Although Bugs2000 is quite widely used in the palaeoentomological community it is very rarely cited, with the exception of the developers and their immediate colleagues. This is not an uncommon situation for databases, and it appears that authors are not always aware that they should cite database software, and perhaps regard it simply as a passive tools in their work. The latter is far from the truth, as the use of Bugs, as opposed to literature or other data sources, directly influences the interpretations made by way of the extent of its data and tools. Citing the correct data source and version is essential for the reproducibility of scientific results, as much for fossil beetle work as for any field. For example, conclusions drawn from DNA analyses using a demographic database on the population of Iceland could be different to those drawn from a database on the UK population, the importance of citing the correct data source being obvious in this example. Computer software is afforded the same copyright status as scientific papers, and it is thus not only ethical to cite their use but also a legal requirement. Databases also have the same status as any archive, and must be cited as a data source when used. Unfortunately the citation indices are yet to cater to data sources, and it is currently not possible to track the degree of citation of the Bugs software. The publication of this thesis may help the situation to a degree, but the authors recommend using the citation form suggested in the software interface, which always reports the versions of the program components and database being used.

2.8 A Brief Developmental History of Bugs

The Bugs project essentially began c. 1989 with the development of the first Bugs database (Sadler *et al.*, 1992). The software was primarily designed to reduce the amount of time required to look up ecology and habitat data for beetle species found in samples.

Previous versions of Bugs have been described elsewhere (Buckland, 2000), and they are summarised below along with their primary references. The intention here is to give an overview of the long term nature of the semi-professional development of a research tool, and to illustrate the tremendous amount of work that has gone into the Bugs project. The author of this thesis has been involved in the project since the early 1990s, and took over the software development side in 1996.

2.8.1 Design criteria for Bugs

The following criteria formed part of the Bugs initiative from the very beginning.

1. It should run on standard home PC, either standalone or through commonly available software.
2. It should be easy to distribute, and independent of installed database software.
3. It should be user friendly and aimed at a poorly computer literate user base.

The following criteria were arrived at during the development of BugsCEP and its components, and only became of real importance in the latest version:

4. It should be expandable and updatable – in terms of both the data and the program.
5. All calibration data should be visible and methods transparent.

2.8.2 Implementing the design criteria

Although the design criteria may appear relatively simple, their implementation has not always been straight forward.

1. Bugs has grown enormously since inception – the original was only about 8 MB, and BugsCEP is between 50 MB and 150 MB depending on installation. Storage space has at various stages been a concern, but technology has fortunately progressed faster than the development of the database and this has ceased to be a major concern. The amount of data in BugsCEP is no longer something that would be considered particularly large, and most current home PC's can run BugsCEP without significant delays when processing.
2. Although BugsCEP is dependent on the MS Access database management system, the developer software includes that facility to create runtime packages for distribution. That is to say that MS Office Developer Edition includes the tools necessary to produce database orientated software that can be distributed to computers that do not have the Access database software installed.
3. This was relatively difficult with the original MS-DOS version of Bugs due to the inflexibility of the dbase4 – Clipper environment (Sadler *et al.*, 1992). With the release of MS Windows, creating GUIs (Graphical User Interfaces) became increasingly easy, and subsequent releases of developer environments such as MS Office Developers Edition have only added to this. A major disadvantage of the MS-DOS version of Bugs was the need for the user to enter search criteria as exact SQL 'where' statements. Although this allowed for accurate retrieval it was not ideal in that the majority of users were unfamiliar with SQL, and even those that were had trouble remembering where to put brackets and commas. Later versions replaced this system with a more intuitive interface where users could enter search terms for biology, distribution and RDB, and select the appropriate search logic from drop down boxes. BugsCEP takes this a step further by providing a comprehensive search system covering the majority of data areas. Much of the usability of recent versions of Bugs is

thanks to the development of object orientated programming, and the relative ease of creating user interfaces in the MS Windows environment, through developer tools which have implemented this.

4. Updates have always been provided for download when significant changes to the data or program have been implemented. Data updates have proved problematic only where users have entered a large amount of data themselves. Unfortunately the system does not include the ability to update an existing copy of the data or program on a user's computer, only replace it. This means that any additions users have made will be overwritten when the update is installed. The separation of Bugs into frontend and backend (program and data) files reduces the problem to great extent by allowing the program part to be updated without overwriting the user's data. The data update problem still remains to be solved however, and at the moment it can only be advised that users export their sites before updating. Bugs was not designed with the end user addition of species data in mind, although it is understood that users with particular specialisms could find this necessary. The current strategy is to keep species and bibliographic data entry centralized with the developers. Should BugsCEP become widely used, this strategy may have to be reviewed with respect to a clearing house or remote web based data entry facility, although such options would require considerably more management than the current system.
5. This criterion was arrived at due to an awareness of the increasing mathematical complexity, and reducing transparency of palaeoenvironmental reconstruction methods. The comment by a conference poster presenter, whose results had proved to be statistically insignificant: '...we hope that more advanced statistical methods will find a significant relationship' suitably summarising the trend. An acceptance that more complex methods will provide better results is a little naive, and at odds with Occam's metaphorical razor. In addition, the assumption that smaller error estimates always mean a better reconstruction is erroneous, and the complexity of the data manipulation often hides many unquantified errors. It was thus the intent of the author to keep any statistical manipulations provided by BugsCEP simple and effective. Beetle abundance data are subject to many taphonomic problems before the final dataset is arrived at. These problems are difficult, if not impossible to quantify (see Chapter 4) – and thus have errors that are difficult or impossible to assess. More complex statistics applied to these datasets often complicate these errors even further. BugsMCR fails very slightly in fulfilling this aim in that it is currently not possible to directly examine the thermal envelope of a species within the program, although the user can create samples with single species and examine the envelopes in MS Excel. This is an unfortunate consequence of prioritization of features, and will be included in a future version.

2.8.3 Version history

Bugs

Primary reference: Sadler *et al.* (1992)

The original program was developed by Jon Sadler as part of his PhD at Sheffield University, UK, with assistance from Mike Rains and Paul Buckland. It was constructed in dBase and compiled with Clipper, and ran in MS-DOS. A manual was produced by Philip Buckland in the early 1990s.

Bugs v.2

Unpublished

Version 2 was developed in VisualBASIC as a standalone MS Windows application, but dropped in the early prototype stages due to lack of time and funding. This was the first Windows version of Bugs.

FoxBugs

Unpublished Masters Thesis, Sheffield University, UK

FoxBugs was developed in the Borland FoxPro DBMS as part of an MSc project in computer science. A prototype system was completed, but further development was dropped when Sheffield University withdrew support for FoxPro, and switched to Microsoft products. Ironically enough Microsoft later acquired FoxPro.

Bugs for Access 2

Primary references: Yuan Zhou (*Unpubl.*); Buckland *et al.* (1997)

The first incarnation of the present series of Bugs programs developed in MS Access. The majority of programming was undertaken by Yuan Zhuo Don, with assistance and later refinements from Philip Buckland. Yuan Zhou wrote Bugs as part of his Computer Science MSc (1995) at Sheffield University, UK, and stayed with the project until 1996.

Bugs for Access 97

Unpublished (Philip Buckland)

The Bugs2000 interface began to take shape in the MS Access 97 RDBMS (which superseded Access 95 so quickly that few people noticed), with a number of improvements in interface design over the Access 2 version.

Bugs2000

Primary reference: Buckland (2000), website <http://www.bugs2000.org>

Bugs2000 was developed in MS Access 2000 and Visual Basic for Applications (VBA), and went through several versions leading up to the most recent v.5. Most of the releases prior to v.5 were bug fixes of various forms, the most common problems arising from variations in MS Windows installations and different versions of Windows. Data were added continually by Paul Buckland, and programming was initially by the author of this thesis, around some of Yuan Zhuo's original core, converted from Access 2. Additional maintenance and restructuring work was carried out by a group of MSc students under the umbrella of Sheffield University's 'Genesys Solutions'^{xiv} software development house, as part of their joint MSc computer science project. The Genesys team undertook the following work:

- split the database into frontend-backend, including facilities for relocation of the backend
- a degree of structural normalisation (some of which was subsequently reversed in BugsCEP, see section 2.4.3)
- fixed a number of errors including those associated with non-English characters in references
- reconstructed the countsheet system to work from externally stored MS Excel files
- implemented a degree of user level security

BugsCEP – Bugs Coleopteran Ecology Package

Primary reference: This thesis, website <http://www.bugscep.com/>

BugsCEP is the version described in this thesis, developed in the MS Office 2000 Developer environment, using MS Access and VBA, in addition to MS Excel for a number of graphing routines and export functions. BugsCEP is a completely new piece of software built around an entirely new,

^{xiv} <http://www.genesys.shef.ac.uk/>

more efficient database structure, as described below. The database design, programming and conversion of data from the previous version was undertaken by the author of this thesis. Paul Buckland has been responsible for the majority of the data entry.

2.8.4 Bugs spin-offs

The concept of Bugs inevitably sparked interest from researchers in other branches of Quaternary science, and the development of version for other proxies were considered.

2.8.4.1 Slugs – Molluscan Database

Primary reference: Unpublished, but mentioned in articles as given below.

A prototype molluscan database was created from the Bugs (v.1. MS-DOS) (Sadler *et al.*, 1992) and subsequently Bugs97 systems (v.2, Figure 2.6), and a small amount of data entered to demonstrate the possibilities. Slugs remains suspended and unsupported, but is available should anyone be interested in taking up development.

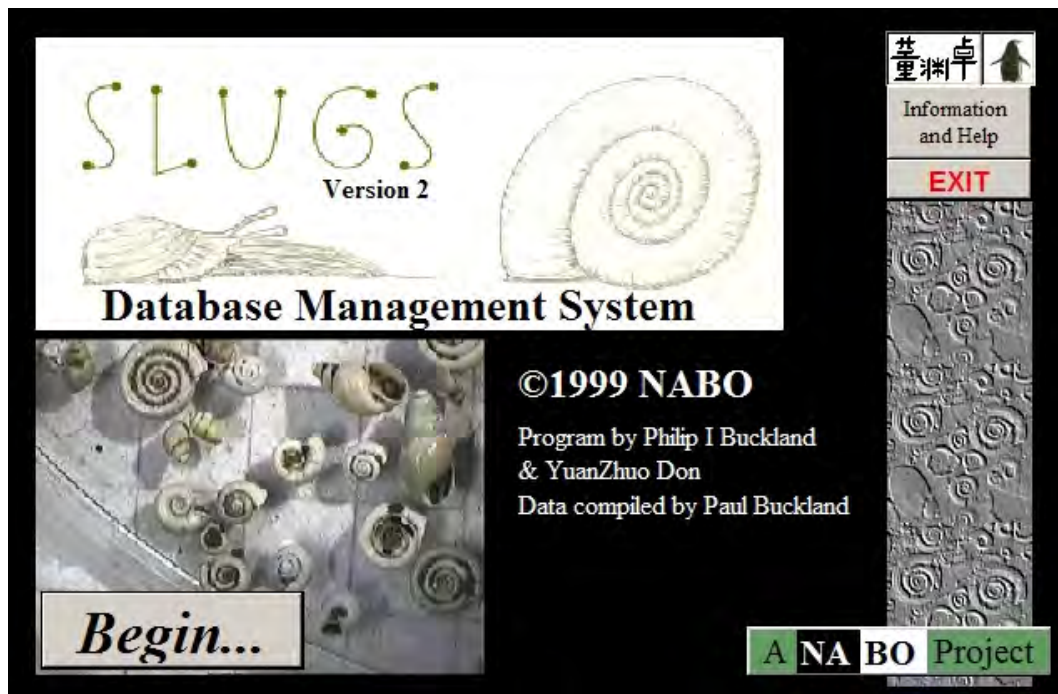


Figure 2.6. Start-up screen of Slugs v.2 prototype molluscan database

2.8.4.2 EgBugs – The Egyptian Coleoptera Database

Primary reference: Buckland *et al.* (*in press*)

EgBugs (Figure 2.7) was developed from the Bugs97 system with an aim towards providing a research and teaching tool for use in the North African and Eastern Mediterranean regions. Its emphasis was on species of use in archaeological investigations in these regions, with the initial scope centred around Egyptian archaeology (hence the name). Programming was by the author of this thesis and Yuan Zhuo Don, and data entry, including the adaptation of the Central European taxonomic code to Egyptian species, was by Paul Buckland and Eva Panagiotakopulu.

The environmental and geographical diversity of beetles makes them ideal for studying environmental change, but also means that attempting to make databases which encompass larger geographical scope will encounter problems. As the database is built around a standard series of codes developed by Koch

(1989) for the Central European beetle fauna, there are problems when the system is applied to other areas, and an international standardization of taxonomic codes would help to resolve the problem of uniquely identifying each taxon numerically. In addition, there is the problem that the habitat preference of species may be geographically dependent. *Sitophilus granarius* L., the grain weevil, for example, is entirely synanthropic in northern Europe, where it only survives in habitats created by humans, whereas it is thought that the species' natural habitat in the Middle East may be in nests of seed storing rodents (Buckland, 1990). A database orientated system, which utilized codes for species ecology encompassing both these regions would have to have functions built in to enable users to limit the scope of fossil habitat related investigations by region. This would avoid the problem of projecting the natural Middle Eastern habitat onto the prehistory of Northern Europe, where it is infeasible. An alternative, as chosen in the current scope of the Bugs project, is to limit the geographical extent of the database to a region where assumptions of habitat specificity are reasonably certain. There is inevitably an element of circular reasoning in doing this, and where long timescales are involved the danger of excluding possible interpretations on the grounds of geographical improbability does present potential problems. An understanding of Coleopteran ecology and biogeography, along with the fundamentals of palaeoecology and Quaternary geology are thus essential when interpreting the results of any fossil insect fauna.

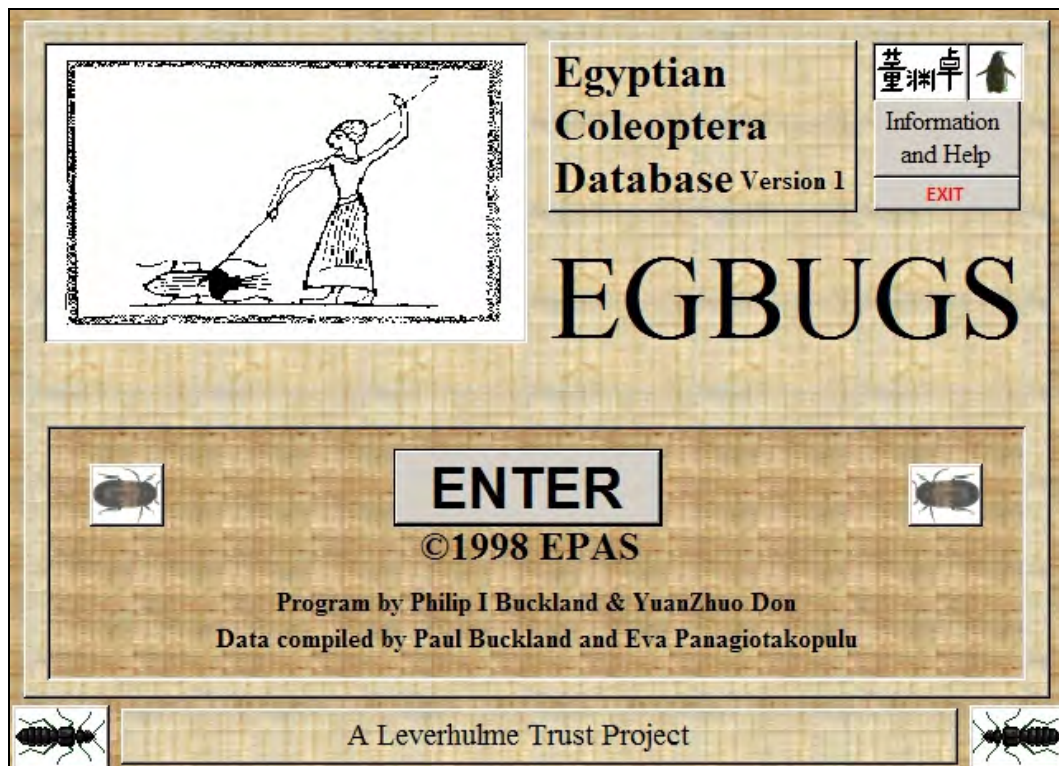


Figure 2.7. Start-up screen of EgBugs – the Egyptian Entomology Database.

2.8.4.3 Other embryonic versions

A plant macrofossil version was investigated, but dropped due to lack of momentum and the need for more structural changes to incorporate ethnographic data and plant use information, along with other modern reference data of interest to archaeologists. A preliminary database structure has been sketched by the author of this thesis, Roger Engelmark and Karin Viklund at the Environmental Archaeology Lab in Umeå, Sweden, and forms part of the SEAD project (see section 1.2.4) prototype multi-proxy database for environmental archaeology (Buckland *et al.*, 2006).

2.8.5 Other Related Software

2.8.5.1 MS-DOS MCR Software – RECON & RECON2 and related programs

Primary references: Atkinson *et al.* (1986), Perry (1986).

Although the original MS-DOS based Mutual Climatic Range software was never formally published or distributed^{xv}, it constituted a major advance in the computerization of palaeoentomology. The two versions, commonly referred to as MCRBirm (Birmingham) and MCR UEA (University of East Anglia) after their respective origins, provided computerized climate reconstructions from beetle assemblages. The reconstruction programs were called RECON and RECON2 respectively, the Birmingham system being the first to be developed. The packages also included a number of tools for data compilation, calculation and display, some of which were restricted to mainframe terminals. Despite the limited user-friendliness of the MS-DOS environment, this was a significant improvement over the use of overlain transparency sheets for climate reconstruction. It was not only faster, but more accurate in terms of removing a number of significant human error factors from the calculation process. Both reconstruction programs ran in MS-DOS, and were compiled FORTRAN programs. The author of this thesis has not been able to run RECON, due to compatibility issues, but was able to extract the thermal envelope data for use in BugsCEP as described below.

2.8.5.2 BugsMCR (standalone version)

Primary references: This thesis, and design outlined in Buckland & Buckland (2002)

BugsMCR is the name of the Bugs and Windows implementation of the MCR method (Atkinson *et al.*, 1986), and was used as the name of the standalone prototype software which is now integrated in BugsCEP. It went through several incarnations before the final prototype was arrived at, and was made available for download from the Bugs2000 website.

The thermal envelope data from the MCRBirm (RECON) was converted and imported into an MS Access 2000 database, and forms the backbone of BugsMCR. Intuitive graphical user interfaces, and export and graphing functions provided a significantly improvement in usability over the previous MS-DOS MCR implementations. A number of more advanced calculation tools were developed within the program, including an application of jackknifing, and rudimentary temperature based prediction (included in BugsCEP). The program was successfully used in teaching in Umeå (Sweden), Bournemouth (UK) and Royal Holloway (UK), and was tested in a number of other departments.

2.8.5.3 BugStats (standalone version)

Primary references: This thesis, and used in producing results for Buckland (2005)

BugStats (previously spelt BugsStats) is the name of the environmental reconstruction component of BugsCEP, and was also the name of the standalone prototype software. BugStats provides semi-quantitative graphical habitat reconstructions from fossil assemblages (although modern studies may also have use for it). It is particularly useful for illustrating changes in habitats over time, and was built primarily for helping in the interpretation of Quaternary sequences. A number of calculation and output options have been built into the current version, including correlation coefficients.

^{xv} The PC software was available for free on request.

3 BugsCEP Database System Description

BugsCEP, the welcome screen of which is shown in Figure 3.1, is a complicated piece of software, consisting of over 30 000 lines of program code in addition to the c. 200 000 individual data records. The *data* and the *program* together constitute an advanced research and teaching tool far more useful than the data alone could possibly be. The data and program are described in this chapter, with minimal attention paid to the applications of the system, some of which are described through case studies in Chapter 6. Note that the descriptions are orientated more towards users than developers, and no attempt is made to follow any particular system analytical schemesⁱ. Some basic entity-relationship models – conceptual diagrams that explain aspects of data and program flow – are included where their use is more practical than written descriptions alone. The second half of this chapter, from section 3.4 onwards, functions as a descriptive manual, and may be used in combination with the inline help when using the program. Section 3.5 provides examples of the reports described in the preceding sections, along with basic instructions on how to create them.

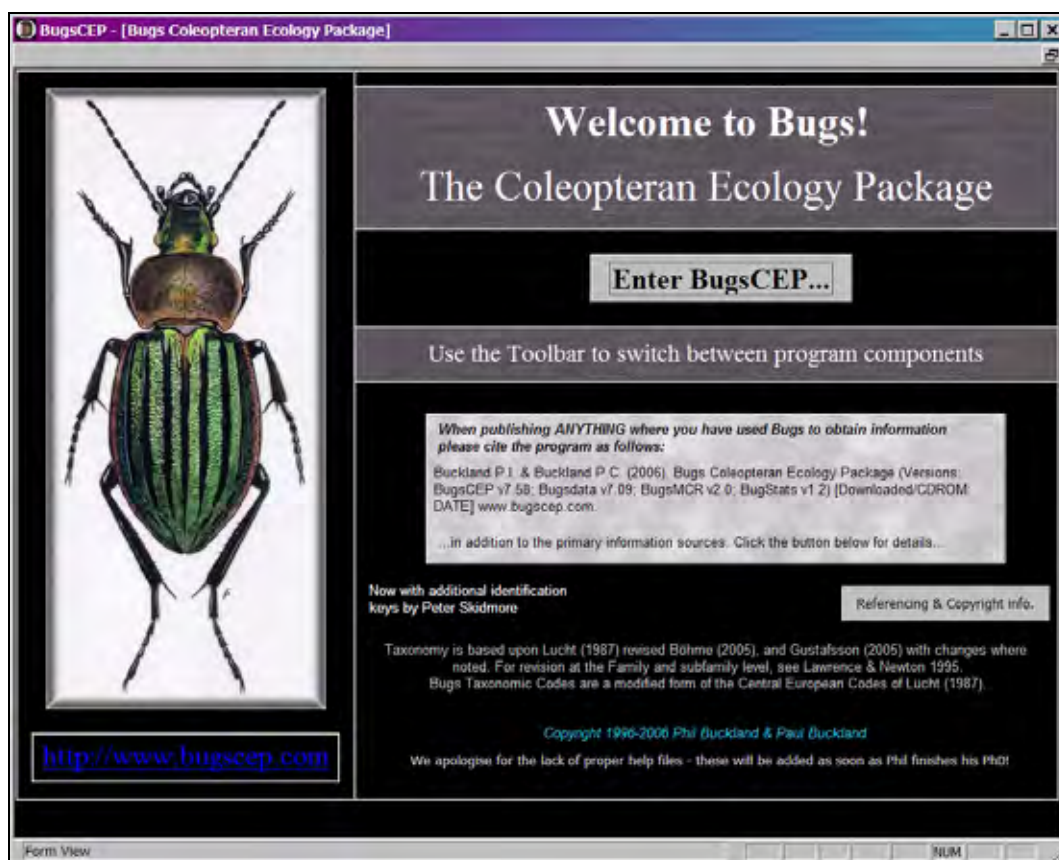


Figure 3.1. Screenshot: BugsCEP start-up screen.

3.1 Data Area Descriptions

The data in BugsCEP can, for the sake of practicality, be divided into three areas: common data, modern reference/calibration data, and fossil/site related data. This is illustrated in the database structure diagram (Figure 2.3), which also shows how the tables mentioned below are related. Data are distributed among the 43 tables in a logic that minimises the amount of duplication of data and empty cells, and groups of tables are used together to store combinations of data items that can be extracted as information. These groups are described below, using the major data area divisions outlined above,

ⁱ There are innumerable software development schemes to choose from, search the Internet for tips.

and with an aim towards providing an insight into the way data are organised with respect to the information accessible via the user interfaces, as described in section 3.4 and onwards.

3.1.1 Common data area

This area contains data that are of relevance to both modern and fossil specimens, or of a structural nature. The master taxonomic index, identification keys, synonyms, taxonomic notes and bibliography form the bulk of this data.

3.1.1.1 Master species list and taxonomic code

The *INDEX* table is a nodal element in the BugsCEP database structure, and stores the master taxonomic list to which all other taxon data are in some way linked. The CODE field – the primary key for this table – is used to identify taxa uniquely, and changes in this cascade to all related tables. This means that only the taxonomic CODEs, and not full species names, need to be stored in any of the related tables. This enables the developer to easily see which taxon is attached to which data throughout the database when debugging or constructing queries. The taxonomy is based on Lucht (1987), as revised in Böhme (2005) and Gustafsson (2005), although some major revisions, such as the repositioning of Dytiscidae before Carabidae as followed by Gustafsson 2005, have not yet been implemented.

The form of the taxonomic code remains the same as that used in the original Bugs database (Sadler *et al.*, 1992), and is derived from the Central European Code of Lucht (1987) as follows:

CODE: 4.0310040 =

4.	031	004	0
Family	Genus	Species	‘Flag’

This example:
 Dytiscidae *Dytiscus* *marginalis* -

The ‘flag’ is an extra digit that allows for the insertion of species not in the Central European list, and can be used to cope with a limited amount of taxonomic readjustment. The use of fractional numbers allows for the relatively easy addition of further ‘flag’ digits should it be needed, although some reprogramming and interface adjustments would be needed to handle this in the software. Note that the authority, L. in the above example, is not a component of the taxonomic code.

The gymnastics necessary around the taxonomic code in order to encompass other regional faunas are evident in the Egyptian database (Buckland *et al.*, *in press*). This situation will have to be resolved in subsequent versions, and it may be necessary to include support for parallel taxonomic systems to provide greater international support.

3.1.1.2 Synonyms

As discussed earlier, insect taxonomy is complex, and undergoes constant revision. This leads to the fact that the same species may have been described under different names by different authors and subsequently synonymised. BugsCEP caters for this by allowing the storage of any number of synonyms for a taxon, along with supporting references and notes. These synonyms are searchable in

ⁱⁱ See Chapter 3, on the development of BugsCEP’s database structure.

various areas of the program, and can be used as an alternative navigation method on the main screen (3.4.1.4).

3.1.1.3 Taxonomic notes

Notes on taxonomy, such as those relating to identification or the splitting of species, are stored as paired reference-data couplets, where the reference always gives the source of the data.

3.1.1.4 Measurable attributes

Measurable attributes are stored in a table designed for flexibility which can be adapted to any quantifiable trait. At the time of writing, this table is only used to store body length dataⁱⁱⁱ in the format shown in Table 3.1.

Table 3.1. Example of size data stored in the measurable attributes table.

CODE	Attribute Type	Attribute Measure	Value	Units
01.00100200	Length	Min	15	mm
01.00100200	Length	Max	19	mm

The example shows the minimum and maximum recorded body lengths of the tiger beetle *Cicindela sylvatica* L., measured in millimetres. At the moment there are no constraints to the entry of attributes – as long as the Value field contains a number – and users are free to enter any useful data. As any number of quantifiable attributes can be stored for each taxon, this feature has great potential for morphometric studies, and even environmental studies if, for example, data such as pollutant tolerance limits were to be included.

3.1.1.5 Identification keys (provided by Peter Skidmore^{iv})

A number of dichotomous identification keys to elements of the British fauna are stored in the *TKeys* table as text (memo) data. These and the taxonomic notes also contain comments on closely related species so far not recorded from the UK, a useful adjunct where fossil material is being considered.

3.1.1.6 Bibliography

The *TBiblio* table contains full references for all information sources cited in BugsCEP. It forms a bibliography of over 3 300 scientific articles and books on modern biology and distribution, fossil record, and to a small extent methodology. About 20 % of the cited works concern palaeo work. Notes may be attached to references, but this feature has not been used much. References are stored as three fields, as shown in Table 3.2

ⁱⁱⁱ These data were extracted by from a number of written and online sources, including Fauna Entomologica Scandinavica and Harde (1998). Additional data were extracted from Skidmore (*unpubl.*).

^{iv} Skidmore (*unpubl.*)

Table 3.2. Bibliography table fields, with example.

Field	Description	Example
Reference (or Ref)	Abbreviated citation used by BugsCEP to link tables, similar to the Harvard System's recommendation for citations in the body of a text. By convention, references to sources on fossil work include the date in brackets, and references to work on modern specimens are without brackets.	Lindroth 1985
Author	Names and initials of all authors and date of publication.	Lindroth, C.H. (1985)
Title	Article/book title, journal, volume, page numbers, publisher etc.	The Carabidae (Coleoptera) of Fennoscandia and Denmark. Fauna Entomologica Scandinavica, 15,1. E.J.Brill, Leiden.

With hindsight, this data should have been stored as individual fields per data item, e.g. Authors; date; title; journal etc., to allow for easier export and more advanced searching. This is, to a large part a legacy problem, in that although a semi-automated conversion would have been possible, it would also have entailed the manual checking of at least 2 500 references imported from the previous version of Bugs. It was decided not to do this, and that the system was efficient enough for its purpose in BugsCEP, with full title searching being sufficiently fast with the small number of records involved.

3.1.2 Modern reference/calibration data

Bugs was initially designed to allow palaeoecologists to analyse their fossil beetle faunas with reference to the modern ecology of the species. The data reflect this, in that the significant bulk of them describes the known biology, distribution and ecology of living insects. Ecological summary data are included in the form of two coded habitat and ecology descriptors – Bugs EcoCodes, and Koch ecology codes (Koch, 1989-92). UK RDB classifications are included for about a quarter of the species. A large number of synonyms are included, which is necessary where the referenced literature is international and of varying age. Thermal envelopes are included for over 400 taxa, enabling at least as good a possibility for climate reconstruction as any previous software. The envelopes have been imported from the existing MCR software RECON (see chapters 2 and 5).

3.1.2.1 Biology and distribution text abstracts

Similar in structure to taxonomic notes above, these are stored as paired reference-data couplets, where the reference always gives the source of the data. If a text abstract refers to another work then the latter work is given its own record, with the data “[See X]” – where ‘X’ is the reference to the secondary referrer. For example, the two records shown in Table 3.3 are amongst the biology data for the dytiscid *Ilybius vittiger* (Gyll.). The later authors have commented on the work of an earlier author. Where the primary data for the species includes an abstract from the primary reference, the “[See X]” data are omitted.

The use of the term ‘biology’ rather than ‘ecology’, denotes the more general nature of the information in the modern data area when compared to the more ecology specific information in the ‘Ecology Summary Codes’ described below. The biology records can vary from broad generalizations to specific local observations. One should *always* take note of the source of the data, and in particular the geographical scope of the work. For example, Nilsson & Holmen (1995) are cited in the example with respect to the “...aquatic Adephaga (Coleoptera) of Fennoscandia and Denmark.”, whereas in the biology data for another dytiscid *Nartus grapii* (Gyll.) are abstracted from Duff (1993) on the “Beetles

of Somerset...”. BugsCEP includes a broad range of specific and general sources, but may be found lacking for certain species and geographical areas, the data for which need expanding.

Table 3.3. Two biology records for the water beetle *Ilybius vittiger* (Gyll.).

CODE	Ref	Data
04.02401010	Nilsson & Holmen 1995	It occurs in small stagnant water-bodies shaded by trees or boulders and often containing Sphagnum mosses. Mainly in wet spruce forests. Often at bog-margins or, especially in the north, in open bogs. The larva was described by Nilsson (1983)
04.02401010	Nilsson 1983a ^v	[See Nilsson & Holmen 1995]

The distribution data are in the form of abstracted records of occurrence. This is a poor substitute for a GIS system with individual find location coordinates, but provides a broad overview of the recent geographical range of species. In addition, there exist a number of national and international systems purposefully built for recording the geographical location of insect and beetle finds. For example: in the UK, MapMate and Recorder^{vi} have been used by many entomologists; Sweden is developing the ArtPortalen^{vii} for centralizing species observations; and internationally, GBIF^{viii} is becoming a strong contender for a global species information centre, including the recording of museum collection data. Currently, none of these systems provide adequate resolution for detailed, pan-European, studies.

Both biology and distribution texts can be searched using the search interface (see 3.4.5).

3.1.2.2 Ecology summary codes (*Bugs EcoCodes* & *Koch ecology codes*)

Two sets of tables hold coded ecological summary values, which can be used to either summarize the ecology of a single taxon or a group of taxa. These tables are fully normalized and allow for any number of codes to be assigned to any taxon. The coding systems – Bugs EcoCodes (created by the author of this thesis and Paul Buckland) and Koch (1989-92) codes are described in detail in Chapter 4.

3.1.2.3 Red Data Book (RDB) classifications

The Red Data Books are an internationally recognised set of publications describing the rarity of species in different countries, and are highly valued in conservation research and planning. RDB status is a measure of the security of a species in a particular country, and the different systems provide classes for a number of degrees of rarity, from ‘extinct’ (EX) to ‘least concern’ (LC), for example, in the IUCN 2001 3.1 system (IUCN, 2001). A flexible set of normalized tables have been put in place to allow for multiple country and international rarity indicators to be stored (Figure 3.2). Although keys are only currently included for the UK, Swedish and international IUCN RDB classifications, it would be easy to add further systems and countries. RDB systems are linked to the bibliography through the ‘Ref’ field in the *TRDBSystems* table, and species are linked to the main *INDEX* table through the ‘CODE’ field in the *TRDB* table.

^v Nilsson (1983) in the references for this thesis.

^{vi} MapMate (2000) © Teknica Ltd - <http://www.mapmate.co.uk>

Recorder 2002 (2004) © NBN Trust - <http://www.nbn.org.uk/recorder>

^{vii} ArtPortalen (2006) (The Species Gateway). Swedish Species Information Centre and Swedish Environmental Protection Agency - <http://www.artportalen.se/>

^{viii} GBIF (2006) <http://www.gbif.org/>

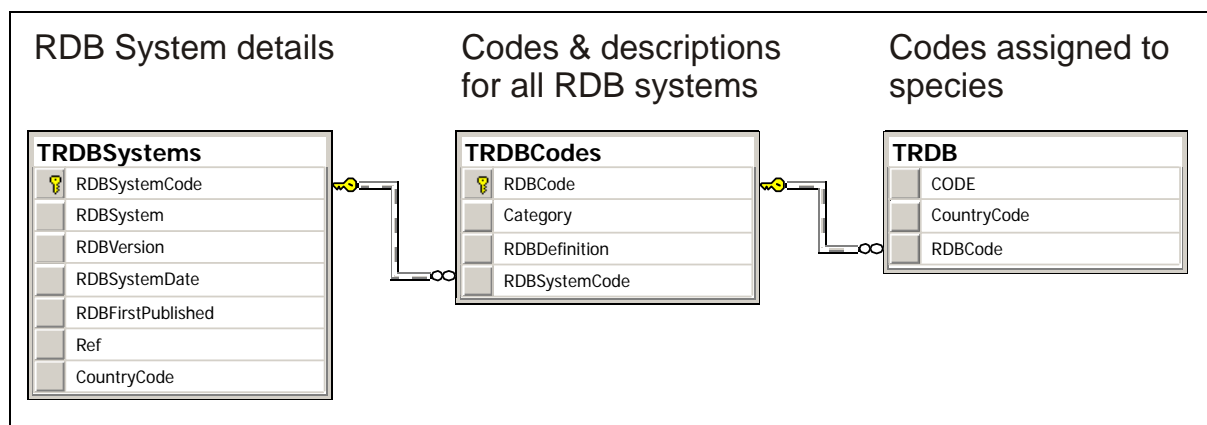


Figure 3.2. Structure of the Red Data Book (RDB) system and code (classification) data area.

3.1.2.4 *Species associations*

Taxa may be associated with any number of other taxa, and the nature of the relationship specified, along with a primary reference stored. This was a late addition to the database, and as such contains very little data – just over 500 records – which reflect the research activities of the developers during the final two years prior to this publication. This should be considered a work in progress, and the ‘Association Types’ available are not formalized at the moment. The field can even be left blank, indicating an unspecified general association, most likely by habitat, although this is not recommended. The majority of association data included so far is from Lindroth (1985, 1986).

3.1.2.5 *Seasons of adult activity*

Data on the seasons during which adults of a species are active are of use to both collectors and palaeoecologists alike, in that it provides a proxy for seasonality, and can be of use when identifying modern specimens. Currently, only UK data for just over 2 200 species are included, but the structure allows for seasonality data for any number of countries to be added^{ix}.

3.1.2.6 *Climate calibration data (thermal envelopes and summaries)*

The calibration *data* will be briefly summarized here with respect to how it is stored in the database. Chapter 5 discusses in detail the BugsMCR system for climate calibration from Coleoptera, which is an implementation of the Mutual Climatic Range (MCR) method (Atkinson *et al.*, 1986).

BugsCEP includes the ‘MCR Birmingham’ data from Russell Coope’s group working in Birmingham (Atkinson *et al.*, 1986; Perry, 1986). To simplify the import process, the original binary grids, which represent one-degree Celsius cells, were imported and converted into a structure where each record represents a one-degree TMax (mean temperature of the warmest month) interval for a taxon. Each record includes 60 fields, which represent the total extent of TRange (difference between the averages of the warmest and coldest months) values for all taxa in one-degree cells. In addition, two further fields allow the records to be assigned to a taxon, and sorted in the correct order.

Information is stored documenting the conversion of each of the taxa in the original MCR dataset into BugsCEP taxa. This information is of no use in BugsCEP, but is stored in the name of backward compatibility and would allow BugsMCR results to be easily compared with those generated by RECON.

^{ix} A minor alteration would also allow the storage of other seasonality data than adult activity – such as larval activity and breeding season.

The use of binary cells allows easy and fast manipulation of the data using SQL. Unfortunately, the use of queries in MS Access leads to file bloating, and in the case of the record and field rich MCR calculations, so much occurred that this convenient method had to be abandoned in favour of VBA calculation. The latter, although it requires more complex programming, is potentially a more flexible and powerful method.

Data for the taxa thermal envelopes, including the outer limits of TMax, TMin and TRange, along with a centre of gravity measure (COG) for each envelope are stored in a summary table. The outer limits are used in the 'BugsMCR Predict' program component (see 3.4.4 and Chapter 5), in which species presences can be predicted from temperature values. The COG values are not currently used by the program, and are part of ongoing development and enhancement of the MCR functionality (see Chapter 5).

3.1.3 Fossil/site data area

BugsCEP has the capability of holding abundance data for both fossil sites – the results of palaeoentomological analyses, and modern sites – the results of entomological collection or survey. For this reason, this data area will be generally referred to as fossil/site data. Sites descriptions can be stored, along with references, the location of specimens, geographical information and other metadata. Samples, site species lists, and abundance or presence/absence data are created and displayed as countsheets, i.e. cross-tabulations of numbers of individuals per species/sample (see Table 1.1).

BugsCEP contains the greater part of the published palaeoentomological data for Europe, and a few extra-European sites, where records, for example of ectoparasites or pests of stored products, are relevant to overall interpretation. In addition to abundance data, an amount of descriptive, geographical and other metadata is stored for sites. There are a few important omissions, such as Coppergate, York (Hall *et al.*, 1992; Kenward & Hall, 1995), where it has not yet proved possible to extract the data in a coherent form^x, and several sites where there are single identifications, sometimes of doubtful validity (e.g. Mjöberg, 1915). Site records, with essential metadata and references have been included for many of these sites, and it should be possible to add them in the near future. All sites are fully referenced, and a simple yes/no indication of the availability of other proxy data for sites is also included. Dating evidence is stored for samples where available, and is divided into calendar, radiometric or period dates to allow for the variable nature of archaeological and Quaternary site dating.

The logical hierarchy of Quaternary, and modern sampled beetle abundance data was introduced in Chapter 1. To an extent, BugsCEP utilizes this logic in its structure, with samples and species lists nested under sites. To improve database efficiency however, spreadsheet data are normalized into a number of tables. As this data area is more complex than the other areas, and indeed, is probably of more interest to the majority of readers, it will be explained here in greater detail, including a detailed enumeration of records in the following section (3.2.1).

3.1.3.1 Sites

A site is defined in BugsCEP as the geographical focal point of an investigation. The vast majority of sites stored in BugsCEP are either the results of palaeoenvironmental or archaeological investigations. Sites contain both the metadata to enable the original sources to be located, and the abundance data relating to the investigation, as described in the following sections.

^x Although the basic species list has been extracted.

3.1.3.2 *Site descriptive data and metadata*

In addition to the site location, in latitude and longitude and national grid systems, BugsCEP can store a limited amount of descriptive information for sites. The aim was not to produce a system specifically for storing site metadata, as there are already systems in existence for this purpose (e.g. ADS – Archaeological Data Service^{xi}), but to provide enough information for the users to be able to get an overview of the current interpretation of the site, locate it geographically, and trace the original work. All of the above could be considered metadata to the species lists, samples, abundances and dates. Site interpretations are necessarily summaries made by past and present authors, and are at some risk of oversimplifying or possibly miss-interpreting the original work. As always, original sources should be consulted for full details.

Site data are stored in the *TSite* table (Figure 3.3), and site references are catered for by an additional linking table, *TSiteRef*, between *TSite* and *TBiblio* (the bibliography), which allows for an unlimited number of references per site. (*TSiteRef* and *TBiblio* have been omitted from Figure 3.3 for clarity, see the full structural diagram, Figure 2.3, for relationship details).

A further table, *TSiteOtherProxies*, is a simple, non-normalized table of yes/no fields to record the presence of other proxy data for a site. This provides a useful overview of the scope of non-beetle analyses at the site. It could potentially be used at a later date as a handle for linking BugsCEP to other proxy databases. References to these other data are progressively being added to the bibliography.

3.1.3.3 *Countsheets: species lists, samples & abundance data*

A countsheet is a collection of species abundance data for a group of samples at a site (the spreadsheet inset of Figure 3.3). A site can have any number of countsheets, and these are displayed and exported by BugsCEP in a spreadsheet like form (see section 3.4.2.3). Data are not actually stored in this form, as a spreadsheet is a highly inefficient structure in that it can potentially contain many blanks, and it is difficult to assign unique identifiers to each abundance value (i.e. spreadsheet cell). Figure 3.3 shows how data and metadata are stored in relation to sites and countsheets. Note that *TFossil* is joined to the central *INDEX* table through the (taxonomic) *CODE* field, which provides the easy display of species names in the spreadsheet display.

Countsheets records are used to store descriptive metadata on the sets of samples, species lists and abundance data for sites, in this way allowing samples to be grouped into meaningful subsets of all those investigated at a site. The metadata also allow countsheets to be defined in terms of *fossil/modern*, and further qualified as *abundance*, *partial abundance*, *presence/absence* or *other* data for statistical purposes.

Samples may be given X, Y and two Z coordinates (top and bottom) to enable easier spatial analyses. The Z coordinates are used to order stratigraphic sequence samples by depth on the dates screen, and will later be available when choosing the sort order for exported spreadsheets and diagrams. As archaeological sites are not often sampled in columns, these data will potentially be sorted in a different manner, and the X, Y coordinates may be more relevant here. An ‘*Other Palaeo*’ category is provided for fossil sites that do not fit either of the above.

^{xi} ADS <http://ads.ahds.ac.uk>, Arts and Humanities Data Service, York, UK.

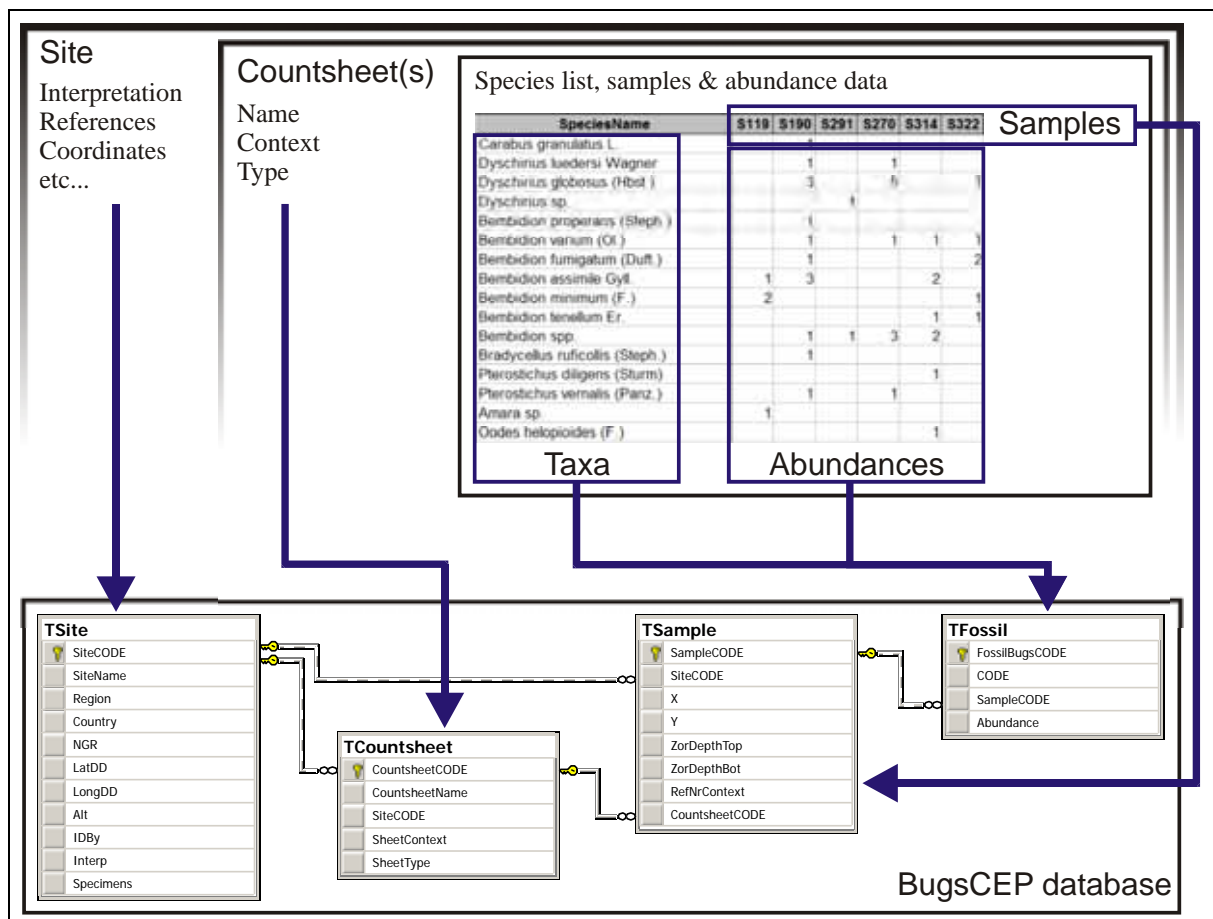


Figure 3.3. Structure of the site and samples area of the database, illustrating how data items are displayed (top) and saved (bottom). Note that the empty cells in the countsheet are not saved in the *TFossil* table.

3.1.3.4 Sample dates

Dates must be attached to samples, and can be defined in terms of three types: calendar, radiometric and period, as described below. Any number of dates can be assigned to a sample, and ranges and uncertainty can be specified using a number of flags as shown in Table 3.4. This allows for a great deal of flexibility in assigning dates, which is especially useful for archaeological sites where dates tend to be limited to contexts containing artefacts or hearths. Previous versions of Bugs only stored dates as lists for sites, and the new version provides much more advanced possibilities for data retrieval and querying.

1. Calendar – dates in specific years or ranges in BP, BC or AD in calendar years. These are most often relevant to archaeological sites, where dates may have been obtained from artefacts or specific events. A somewhat macabre example is that of the single individual of the ‘graveyard beetle’ *Rhizophagus parallellocollis* Gyll. found in the lead coffin of Anne Mowbray^{xii}, which was sealed in 1482 (Stafford, 1971; Girling, *pers. comm.* to Paul Buckland).
2. Radiometric – primarily used for radiocarbon dates, but a number of other radiometric (or similar) dating methods are selectable from the ‘Dating Method’ box. These dates are given in the standard form of *Lab Number* followed by Age BP ± Error, and *always* in radiocarbon (or other method) years.

^{xii} Anne de Mowbray – the tragically short lived 8th Countess of Norfolk, Duchess of York, and Duchess of Norfolk, was buried initially in Westminster Abbey, London in 1482.

3. Period – the least specific dating category, which allows samples to be assigned to a number of age categories. These categories vary in their range from shorter archaeological periods, such as, the 150 year Late Roman UK period from AD 250-400, to considerably longer geological periods such as the Holocene, spanning the last 10 000 years.

Table 3.4. Available sample date range and uncertainty flags.

Type	Flag	Date types
Range	From	Calendar, Period
	To	Calendar, Period
	>	Radiometric
	<	Radiometric
Uncertainty	Ca	Calendar, Period
	?	Calendar, Period
Range with uncertainty	From ca	Calendar, Period
	To ca	Calendar, Period

All dates require that the dating method is chosen from a drop down box, and notes can be added should further clarification be necessary.

3.2 Record Enumeration & Database Size

There are many ways of structuring databases, and so enumerating record counts is often of limited real comparative value between systems. They are, however, useful for comparing the relative amounts of data in different sections of a database if a uniform approach to its construction has been followed. Table 3.5 summarises the record counts for the main areas of BugsCEP’s data. It should be noted that Bugs is updated continually, and that these counts represent the counts at the time of writing of this thesis. There is also a considerable amount of support data, such as look up tables, definition lists etc., not listed here that enable both the user interface and statistical methods to function.

Table 3.5. Summary of BugsCEP data, showing number of records per data area (support tables excluded).

Species related	
5 845	Taxa
4 698	Synonyms
528	Records of species association
1 719	UK RDB rarity records
5 407	Species attribute records (mainly size data)
Bibliographic	
3 373	References
Fossil record and other sample based	
601	Sites
656	Abundance datasets/countsheets
89 917	Fossil record entries (which make up the 656 countsheets)
4 871	Recorded dates, of which: 646 radiometric; 3 685 period; 540 calendar
Climate and Ecology	
436	MCR thermal envelopes
22 884	Biology text excerpts
22 331	Distribution text excerpts
8 302	BugsEcoCode assignments
11 245	Koch Ecology Code assignments
9 726	UK Monthly activity records
Other	
267	Identification keys (by Peter Skidmore)

Earlier papers have presented the physical size of the database in terms of disk space. At the present day these numbers are somewhat irrelevant, due to variations in data compression/compaction methods, database structure (normalization issues in particular), and other software/hardware issues which can cause the same amount of data to occupy different amounts of disk space. The numbers are, in fact, only interesting in terms of nostalgia (cf. older versions) and practical installation requirements; they are presented here in Table 3.6 for the sake of completeness.

Table 3.6. Physical disk space size of BugsCEP database (see text for disclaimer)

BugsCEP program part	18 MB
Bugsdata database part	34 MB

For comparison, the original Bugs software (Sadler *et al.*, 1992) totalled 8 MB, Bugs2000v.4 was 37 MB, and Bugs2000 v.5 (Buckland, 2000) – the version immediately preceding BugsCEP – was

about 150 MB. At around 52 MB, despite a massive increase in data, BugsCEP represents a significant improvement in data storage efficiency over previous versions^{xiii}.

3.2.1 Sites overview

Aside from the fact that BugsCEP is Eurocentric, the geographical distribution of sites primarily reflects the activities of palaeoentomologists, rather than the availability of suitable sediments. The access to cut or eroded peat bogs, frozen lakes^{xiv}, and waterlogged archaeological sites are, however, particularly advantageous for palaeoecology. Naturally, the scope of database also reflects the interests of the developers. Figure 3.4 shows the number of sites per country, and England, as the country where palaeoentomology has been most extensively used and most intensively developed, is clearly the most thoroughly palaeoentomologically investigated.

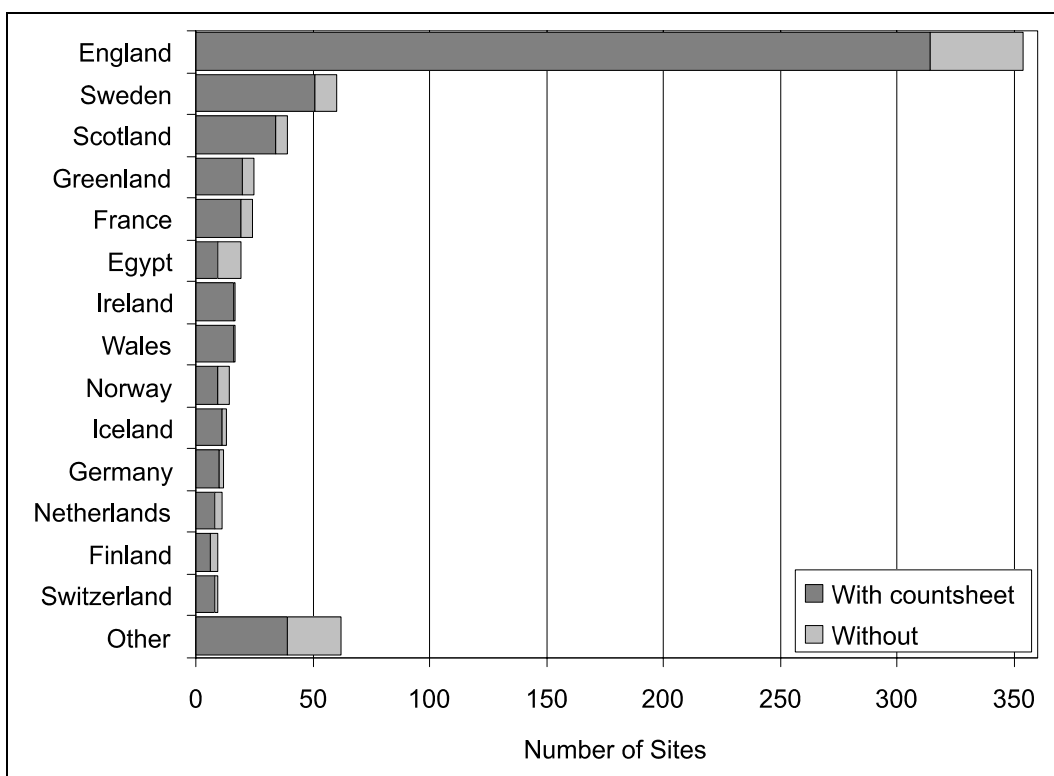


Figure 3.4. Bar graph showing number of sites per country. Lighter parts of the bars represent sites with only summary information and no abundance data. ‘Other’ shows the total sum for all countries with less than nine sites per country.

The majority of sites have been assigned latitude/longitude, and can thus potentially be useful in GIS analyses. Figure 3.5 shows the location of the sites which currently have coordinates, and lists the number of sites per country for which they are lacking. The latter are mainly due to a surprising number of published papers that do not give coordinates for the sites. The map clearly shows the primary concentration of sites in the British Isles.

^{xiii} Note that due to the use of external library files and the MS Access database runtime engine, the actual installation size can be anywhere between 60-150 MB.

^{xiv} Although it can be difficult to obtain sufficient material from the latter.

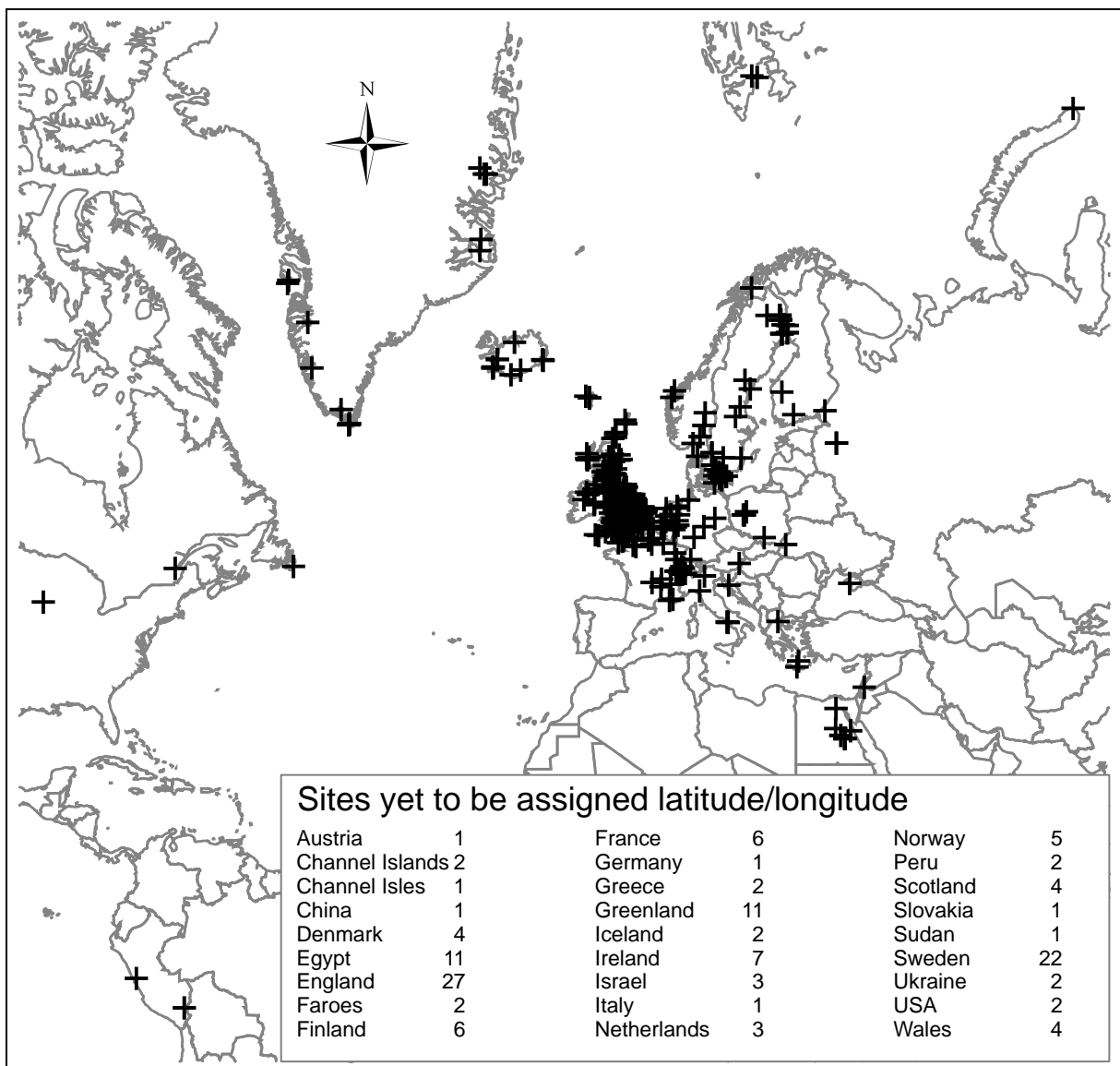


Figure 3.5. Geographical location of sites in the BugsCEP database. Mercator projection.

3.2.2 Countsheets: species lists, samples & abundance data

Table 3.7 shows the number of countsheets stored, at the time of writing, for each of the possible metadata qualifiers which describe the context of the data. Stratigraphic sequences and archaeological datasets make up the bulk of the countsheets, and there is a clear need for the entry of more modern collection data, which would allow for comparative studies between fossil and modern assemblages to be more easily undertaken. The two modern, pitfall trap datasets are discussed in Chapter 6.

Table 3.7. Available countsheet context descriptions and enumeration in BugsCEP.

Timeframe	Context	Count
Fossil	Archaeological contexts	258
	Stratigraphic sequence	345
	Other Palaeo	11
Modern	Pitfall traps	2
	Other Modern	0

The spatial distribution of sites in the core geographical area is shown in Figure 3.6, with stratigraphic sequences and archaeological contexts differentiated (sites without lat/long have been omitted). An interesting point is that Sweden and France have a significantly higher number of sites with stratigraphic sequences (41 and 14 resp.) than archaeological contexts (11 and 4), which most probably reflects the position of palaeoentomology in Quaternary science, rather than archaeology in these countries.

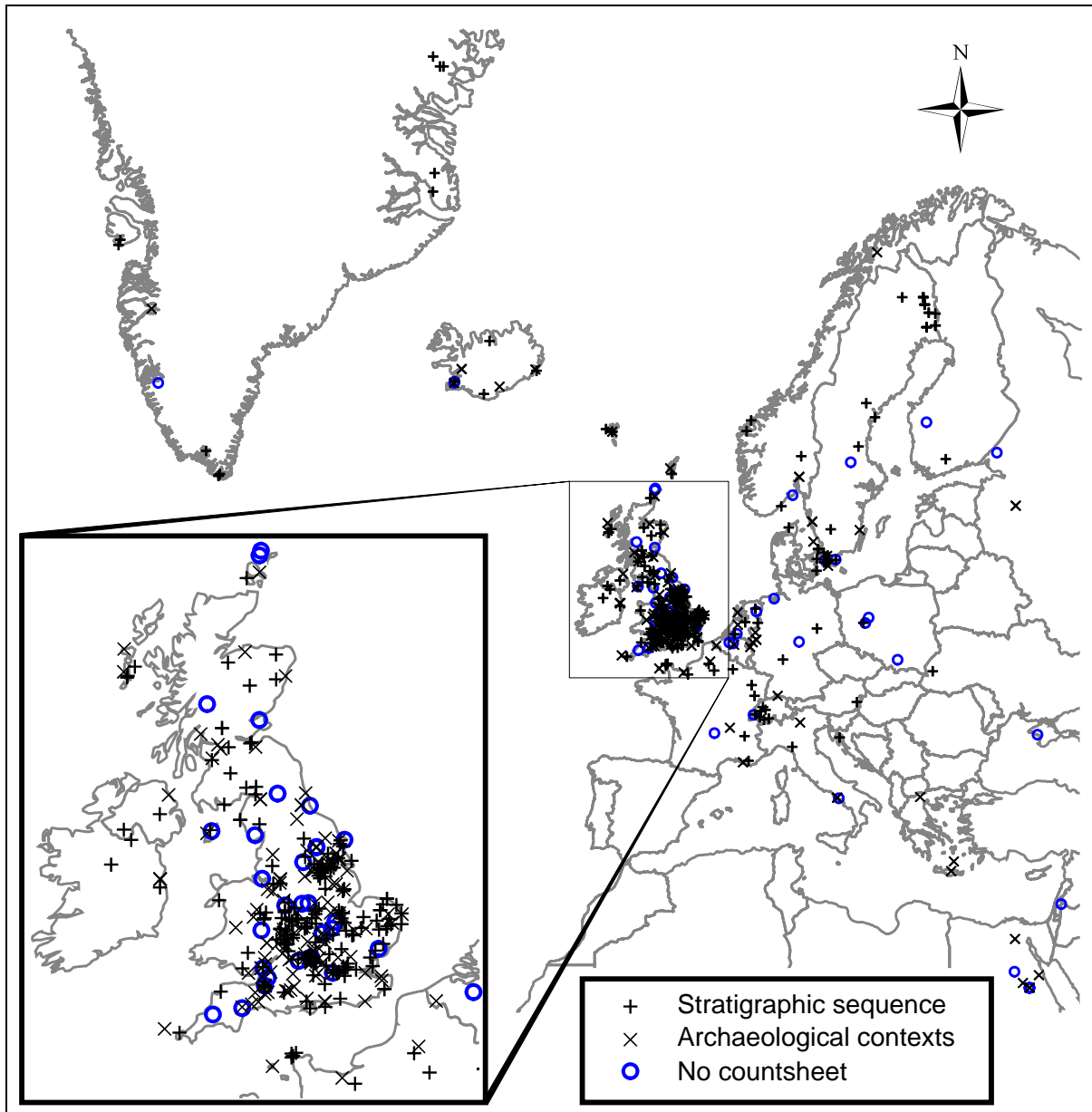


Figure 3.6. Map showing location of sites with stratigraphic sequences, archaeological contexts and no countsheets in the BugsCEP core region.

3.2.3 Sample dates

The enumeration of sample dates is currently a complex matter, as the dating search engine is incomplete, but an approximate summary by dating type is provided here. The final search engine will be able to use a lookup system to search fluidly over all three dating types. Note that a sample may have more than one date of any kind, and that this will affect the histograms below.

3.2.3.1 Radiometric dates

The clear dominance of Holocene and Lateglacial research is indicated in Figure 3.7, although this is of course greatly influenced by the limits of radiocarbon dating. Only 25 dates older than 100 000 BP are currently stored, and have been omitted from the histogram for reasons of clarity. The apparent lack of early-mid Holocene dates is largely a reflection of the poor number of mid-Holocene dated samples (Figure 3.8). Although this may to an extent reflect dating strategies (researchers tend to date the top and bottom of sequences), a cursory examination of the undated samples in BugsCEP suggest that the pattern does indeed reflect a lack of samples as well as dates. This is a significant problem in palaeontomological research, and makes it hard for the existing data to be used to address contemporary research problems such as the structure and development of mid-Holocene forests, and the impact of Palaeolithic and Mesolithic peoples on the landscape (Buckland [*et al.*], 2005). These time slices are a clear target area for future research, and this illustrates the strategic importance of BugsCEP in identifying areas in need of further research.

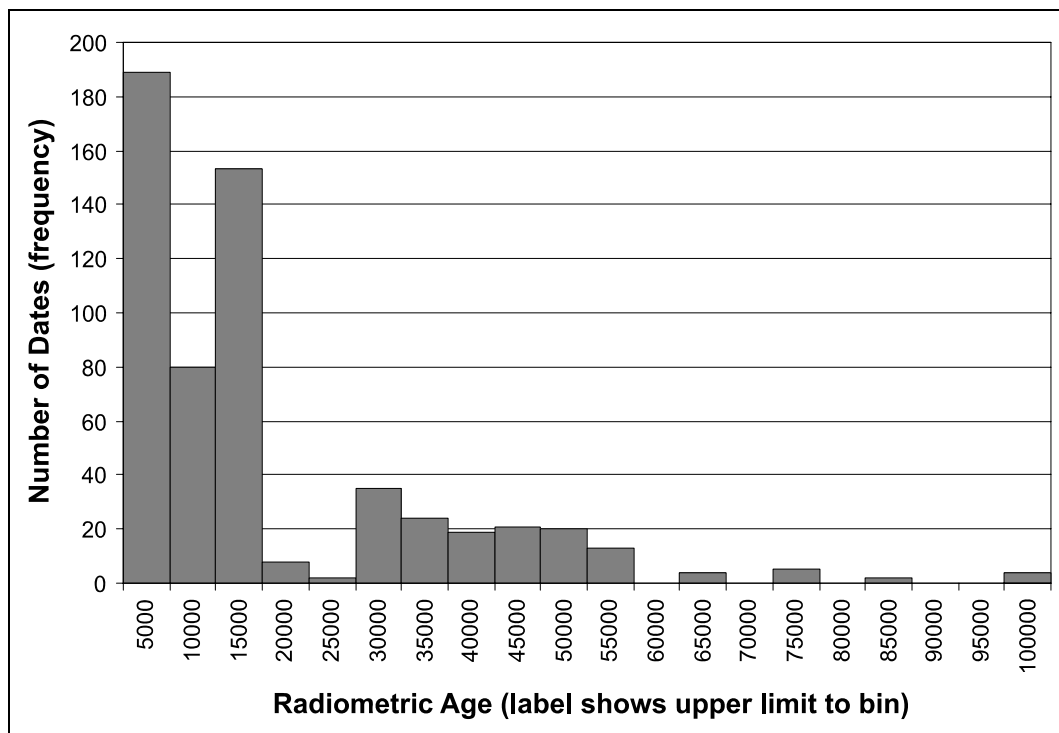


Figure 3.7. Histogram showing number of radiometric dates per 5 000 year period (bin) over the last 100 000 radiometric years.

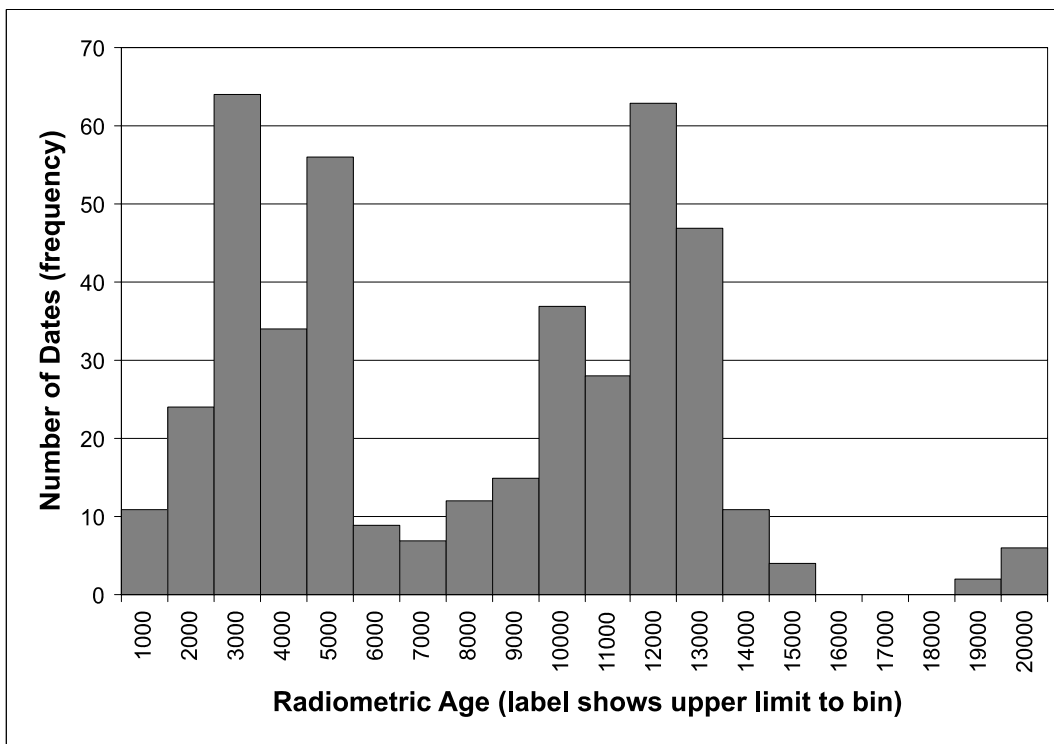


Figure 3.8. Histogram showing number of radiometrically dated dates per 1 000 year period (bin) over the last 20 000 radiometric years.

3.2.3.2 Calendar dates

Calendar dates are predominantly archaeological in nature, and primarily reflect the focus of archaeological investigations in the British Isles on the medieval and Roman periods (Figure 3.9). 53 dates fell beyond the 5 000 year range of this histogram, their spread lending little to the observed pattern and they were thus omitted from the figure to improve resolution.

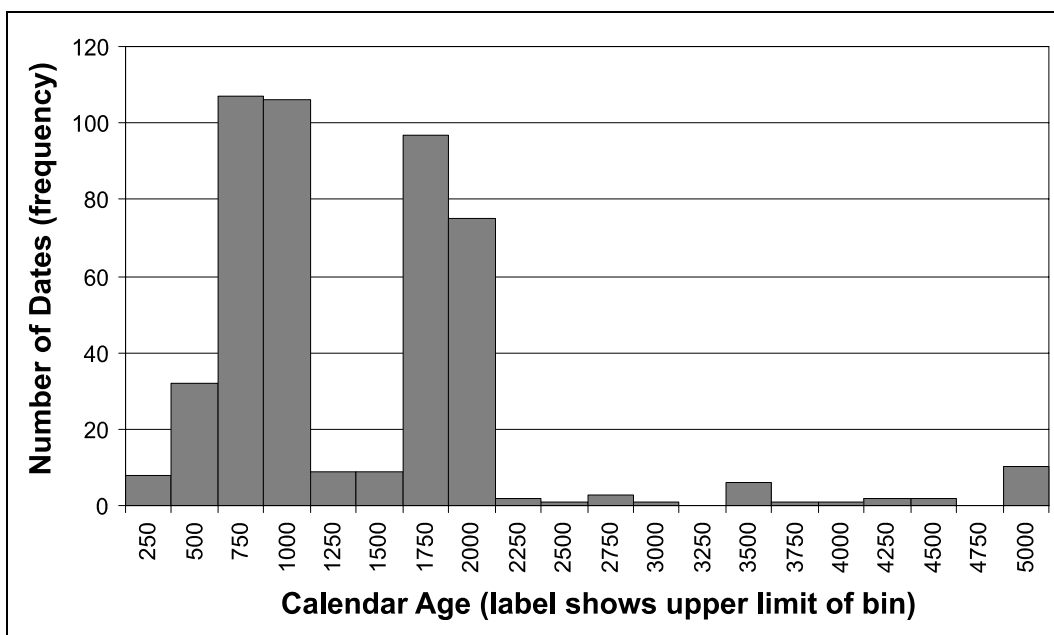


Figure 3.9. Histogram showing number of calendar dated dates per 250 year period over the last 5 000 years (corrected to calendar BP where present = 1950 for limited comparability with radiometric ages).

3.2.3.3 *Period dates*

The period date category is still under development, in that the categories still need to be consolidated and assigned calendar or radiometric equivalents to enable cross searching. The assignment of period dates has been somewhat arbitrary, and most often reflects the context of an investigation, for example, a sample from an archaeological investigation is more likely to be classified as post-medieval than Late Holocene. There are also a number of synonymous periods, such as the Eemian, Ipswichian and (Marine) Oxygen Isotope Stage 5e (Bell & Walker, 2005), and it has been decided to use the name given by the original author to avoid potential misinterpretations.

Period dates have been assigned to a large number of samples that have not formally been dated, where their context gives reason to believe a particular period is appropriate. The primary focus of palaeoentomological research on Holocene sediments is again visible in Figure 3.10.

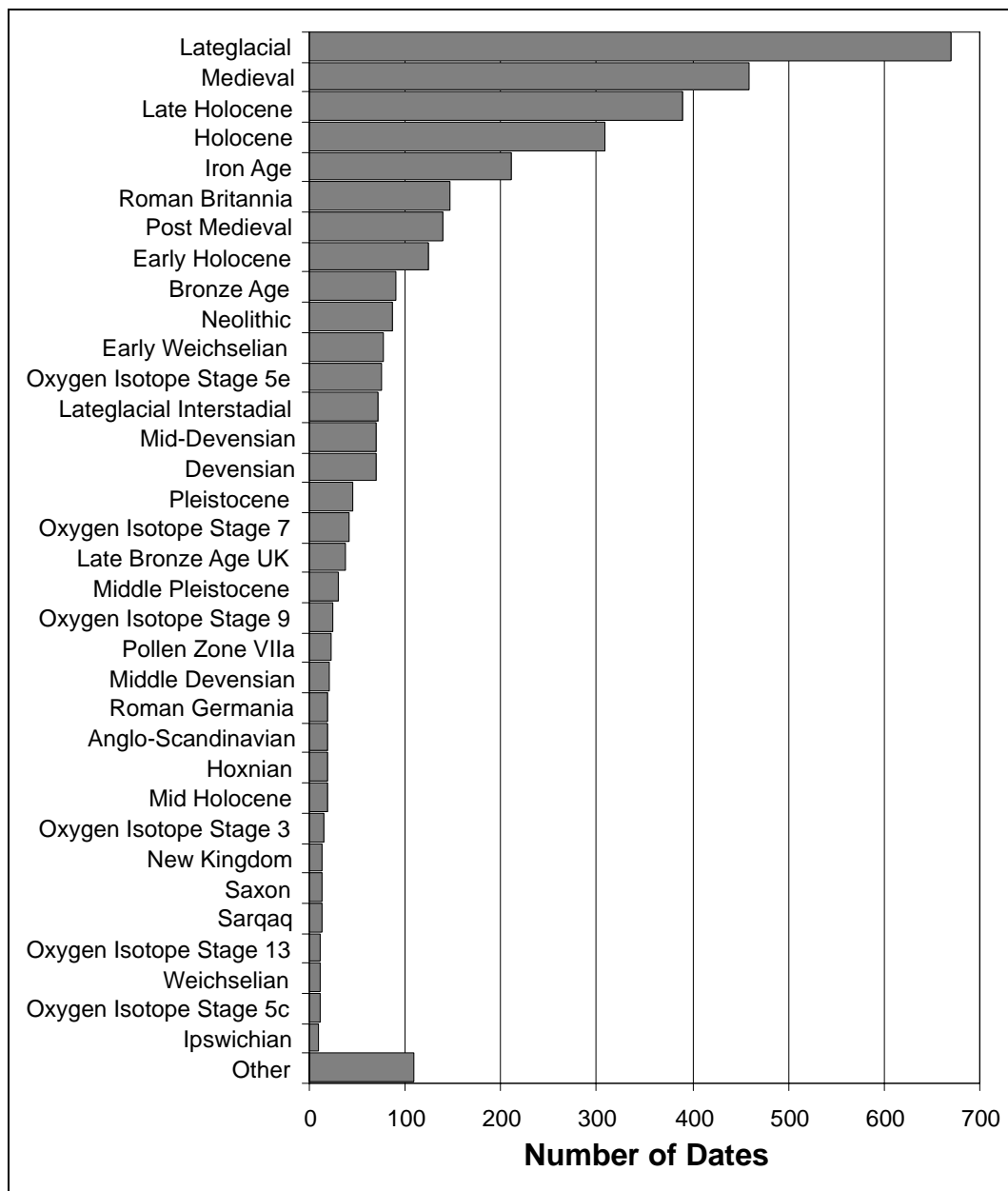


Figure 3.10. Bar chart showing number of dates stored per time period. The 'Other' category is the sum of all periods with less than ten dates.

3.3 User Base

Bugs has never required registration or payment for use, and as such has been freely available for download since 1998, prior to which it was distributed on floppy disks. The tracking of downloads to specific users, a difficult and time consuming task, has been avoided, and as such there is no real record of all users. As mentioned earlier in Chapter 2, there is a general lack of consistency in citing databases and software in scientific publications, despite being covered by the same copyright laws as any other written publication. Quaternary entomology is a small field, with less than 200 researchers world wide, and so the primary user base for the *fossil* side of the database will inevitably be small. However, there has been an increasing amount of interest from modern ecologists, particularly those interested in landscape change or biogeography, and this broadens the potential audience considerably. The Bugs project has taken this interest on board, and enhancements over the previous versions of the software have been implemented into BugsCEP which could increase the usefulness for ecologists considerably.

We have, however, kept track of those users that have specifically requested the database, have requested technical support, or contacted us on other Bugs matters. This information is used here to (anonymously) summarize the proportion of users in different fields and regions, although the numbers and classifications may be approximate due to individuals being active in more than one field. Of the 50 known users, 38 are based at academic institutions, seven are private individuals and five are employed by government institutions including museums. Table 3.8 shows the spread of users by country. Bugs is known to be used in research, teaching and consultancy, and has been implemented as part of teaching modules in the Quaternary in Edinburgh (Scotland), Bournemouth (UK) and Umeå (Sweden). The BugsMCR component has been evaluated for teaching purposes at Royal Holloway (University of London) and East Anglia.

Table 3.8. International distribution of known Bugs users.

Country	Count
UK	33
Sweden	7
Ireland	2
Spain	2
USA	2
Germany	1
Greece	1
Norway	1


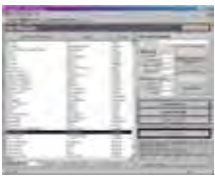
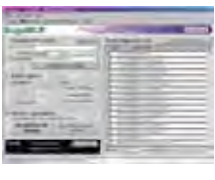


In future versions steps will be made to require user registration so that the use of Bugs can be more accurately approximated. Bugs has previously only been passively marketed, that is to say that there has been a limited amount of promotional material produced, a situation that is not ideal, but is a result of the limited availability of time and funds for marketing. This will hopefully be rectified in the future. This aside, the number of discrete visits to the www.bugscep.com website have steadily increased from around 500 to over 800 per month between November 2006 and March 2007, showing that even passive marketing has some effect.

As an interesting side note, the announcement of BugsCEP on two Coleoptera user groups (british-beetles@yahoogroups.com and coleoptera@yahoogroups.com) was closely followed by 40 downloads. This clearly demonstrates the importance of digital networking and marketing. A similar number of downloads followed the presentation of the software at the Late Neogene Ecosystems Database Workshop (LNED, see Chapter 1), demonstrating the similar importance of making software visible at relevant scientific meetings. None of these downloads are included in the figures above.

3.4 The BugsCEP Program - Detailed Description

BugsCEP can essentially be divided into five interface areas (Table 3.9), which allow the user to undertake a set of specific tasks. An environmental or palaeoecological investigation would inevitably use several, if not all, of these components, but any one can provide independently useful information.

Table 3.9. A summary of the BugsCEP main interface areas. I/O = Input/Output.

1.		<p>Area: BugsCEP Main Screen Primary form: BugsCEP Main Data: Species specific information: ecology, distribution, taxonomy summary of fossil record associations with other species bibliography identification keys Functions: species based data retrieval add/edit species based data add/edit references I/O: export data for selected species export selected references</p>
2.		<p>Area: Site & Countsheet Management Primary form: Site Manager Data: Site specific information: site summaries (interpretation, lat/long, references, etc.) countsheets (species, sample and abundance data) dating evidence Functions: create/edit sites and summaries add/edit species lists, samples and abundance data add/edit sample based dating evidence I/O: import abundance data export abundance data generate site reports</p>
3.		<p>Area: Climate Reconstruction Primary form: BugsMCR Data: Species thermal envelopes and summary values Functions: thermal reconstructions from site abundance data prediction of species presence from temperatures (under development: jackknife calculations) I/O: export reconstruction results export reconstruction graphs export climate space maps and sample species lists</p>
4.		<p>Area: Environmental Reconstruction Primary form: BugStats Data: Coded species habitat summaries Functions: environmental reconstructions from site abundance data correlation coefficients explore habitat codes by species I/O: export reconstruction results export reconstruction graphs generate sample level reconstruction breakdown report export correlation matrices</p>
5.		<p>Area: Bugs Search Explorer Primary form: BugStats Functions: search for species by habitat, distribution, RDB, ecology I/O: export search results (species lists and information) generate report of sites where species present</p>

In addition to the above, there are a number of pop-up and utility forms that are either reused in various areas, such as the Bibliography Browser, or serve maintenance functions. The Bibliography Browser displays references relevant to the database section that it was accessed from, as well as providing a searchable interface to the complete BugsCEP bibliography of over 3 300 books and articles.

3.4.1 BugsCEP Main Screen – basic data retrieval

Upon starting BugsCEP, users are presented with a welcome screen (Figure 3.1) (and on the first run will have to help the program locate the data file and accept the licensing agreement). Clicking [Enter BugsCEP] opens the BugsCEP main screen, which primarily provides access to information on individual species. Initially, this screen shows the seemingly blank screen for the taxon Carabidae indet.^{xv}, but by pressing the [Next Sp. ▼] the user can update the display to show the data for *Cicindela sylvatica* L. (Figure 3.11), which is the first species in the BugsCEP taxonomic order. By default, the main screen shows Biology & Distribution data. This always consists of a reference and data component – where the reference is an abbreviated citation to the data, which itself is a text abstract, or quote from the cited work. The first item of biology data for *C. sylvatica*, for example, describes the species as being found “on sandy ground, especially heathland, but generally local and rare”, with the information being credited to “Harde 1984”^{xvi}.

Clicking on the [Bibliography for panel] button will open the pop-up **Bibliography Browser**^{xvii} and display the full references for all sources cited for the information displayed. The pop-up lists references alphabetically by first author, with “Harde 1984” being the second record in this case (Figure 3.12). The Bibliography Browser allows users to copy individual references with the click of the [Copy Ref.] button, or export a list of all the references for the current species (see section 3.5). The [Show Entire Bibliography] switches display to the full bibliography, and is replaced by a [Show Specific Bibliography] button when clicked. In this way users can easily switch between viewing only the bibliographic information for the current species and the entire BugsCEP bibliography. The browser is also fully searchable, and clicking the [Search...] button opens the ‘Search in references’ dialog form. This can be used to find the first and subsequent occurrences of search terms in the abbreviated references or full titles.

^{xv} Carabidae indet. is the taxon used to label individuals identified only to the level of Carabid family, and as such is taxonomically hierarchically above all Carabid species. The record has therefore no ecological or distribution information itself, as this would be far to generalized to be of any use. It does, however, have a considerable fossil record due to the frequency of fossil specimens that can only be identified to this level.

^{xvi} Harde (1984) in this thesis.

^{xvii} Interface forms will be written in **bold** the first time they are mentioned.

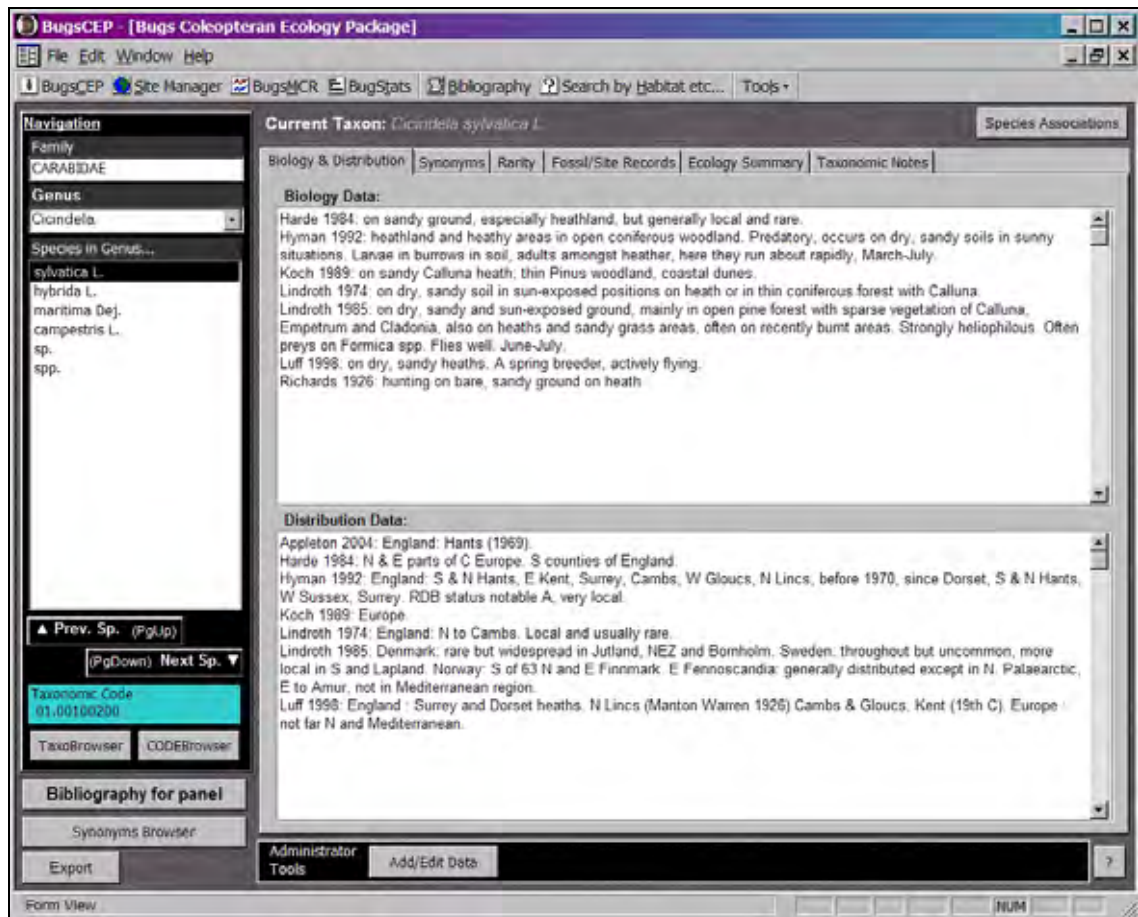


Figure 3.11. Screenshot: BugsCEP main screen showing biology and distribution data for *Cicindela sylvatica* L.

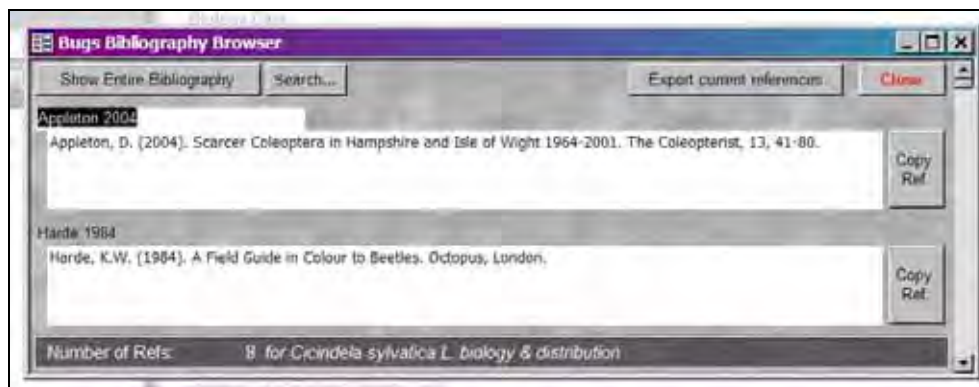


Figure 3.12. Screenshot: Bibliography pop-up showing references for *Cicindela sylvatica* L.

By using paired reference and data items, BugsCEP ensures that there is never any doubt as to the primary source of the information displayed. Also, by using abstracted texts, it removes any possibility of unintentional reinterpretation that would be a problem when summarizing the texts of several authors into a more compact description. This is, of course, always a problem when describing the ecology of species, and is discussed more thoroughly in the context of summary codes later in this thesis (Chapter 4).

The main screen can essentially be divided into five sections, as illustrated in Figure 3.13, which are used to either navigate, display data or activate functions.

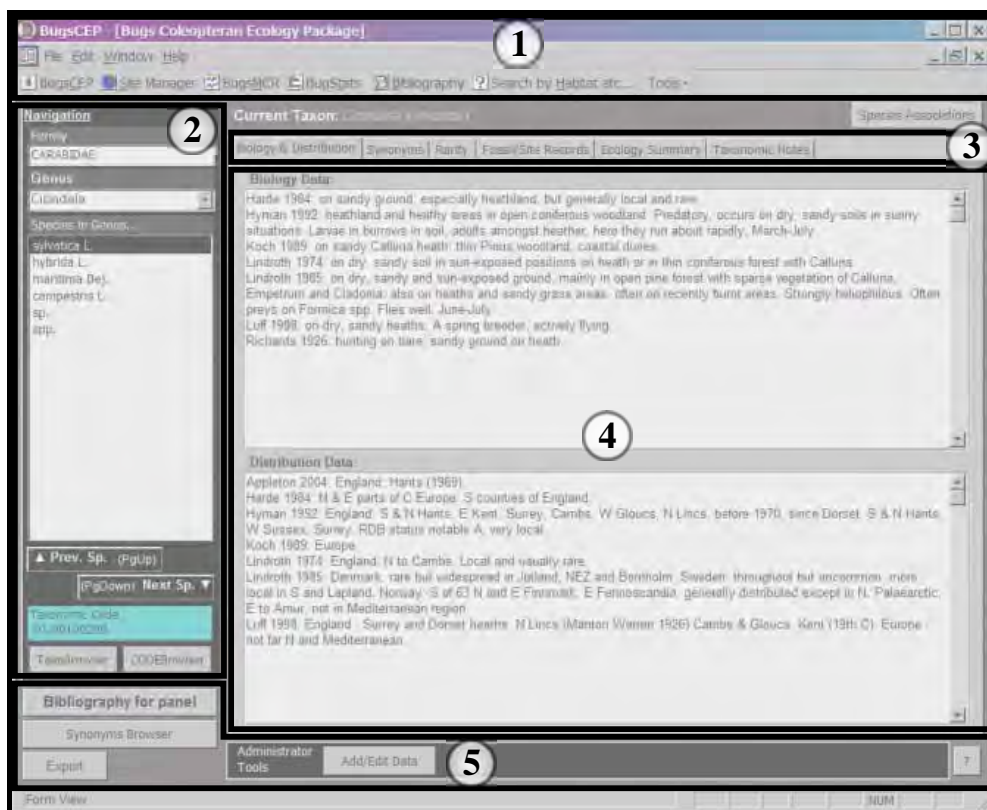


Figure 3.13. BugsCEP Main Screen areas: (1) Title bar, menu bar and component buttons; (2) Navigation panel; (3) Information tabs; (4) Information area; (5) Additional buttons and administrative controls. (The background image is the same as Figure 3.11).

3.4.1.1 Section (1): Title bar, menu bar and component buttons

This area is available from all major interfaces, and allows users to easily switch between program components, and access a number of common functions. The title bar provides the title of the program component, or window, which is currently active. Likewise, the menu items available on the menu bar underneath this will reflect the currently active component, and provide access to functions specific to the latter as well as more general functions available from almost anywhere. Help is available through the menu on most screens, or by pressing the <F1> key on any screen. The Window menu allows users to switch between the program components that are currently open, although this function is unavailable during some critical operations, where its use could lead to loss of data. Standard copy and paste functions are available from the Edit menu^{xviii}. Below this, the BugsCEP toolbar provides access to the major components of the program (see Table 3.9) through a number of buttons.

3.4.1.2 Section (2): Navigation panel

This is specific to the main screen, and allows the user to find any species by a number of means. For most users, the most common navigational path will be to type the required genus into the [Genus]

^{xviii} Please note that the Find and Find Next functions on the Edit menu only allow you to search within text boxes, such as those for biology and distribution. The Search by Habitat function, as described in section 3.4.5, allows users to searching the descriptions of all species.

‘drop down’ box on the left at the top of the panel. BugsCEP will search for the genus as the letters are typed, and pressing the <Enter> key will move to the currently displayed genus. When a genus is selected, the [Species in Genus] list below it will be updated to display all the taxa within that genus, including placeholders for specimens identified to generic level. The first species, taxonomically, in the genus will be highlighted, and the information area (4) will be updated to display data for that species. The [Current Taxon] box just under the toolbar will always show which taxon is currently selected, and the [Taxonomic Code] box will display its numerical equivalent. Note that the [Family] box is not selectable, and always displays the family of the currently selected taxon.

Clicking on the [Prev. Sp.] or [Next Sp.] buttons will move the display to the previous or next taxon respectively in taxonomic order, as will pressing the <Page Up> or <Page Down> keys when the cursor is not in an information box^{xix}.

The Taxonomic Explorer (Figure 3.14) is activated by pressing the [TaxoBrowser] button. This opens a pop-up window which displays three hierarchical panels with all families, genera and species in taxonomic order. Clicking on the [Select Family:] panel will update the [...then Genus:] panel to display the genera within the selected family, which in turn will update the [...then Species] panel to display a list of the taxa within the selected genus. The example shows the navigational path for the Elaterid species *Agriotes pallidulus* (Ill.). Double clicking on the taxon in the final panel, or clicking the [Goto Species...] button will update the main screen to display the available data for the taxon. The pop-up can then either be closed, or moved out of the way so that the main screen is fully visible.

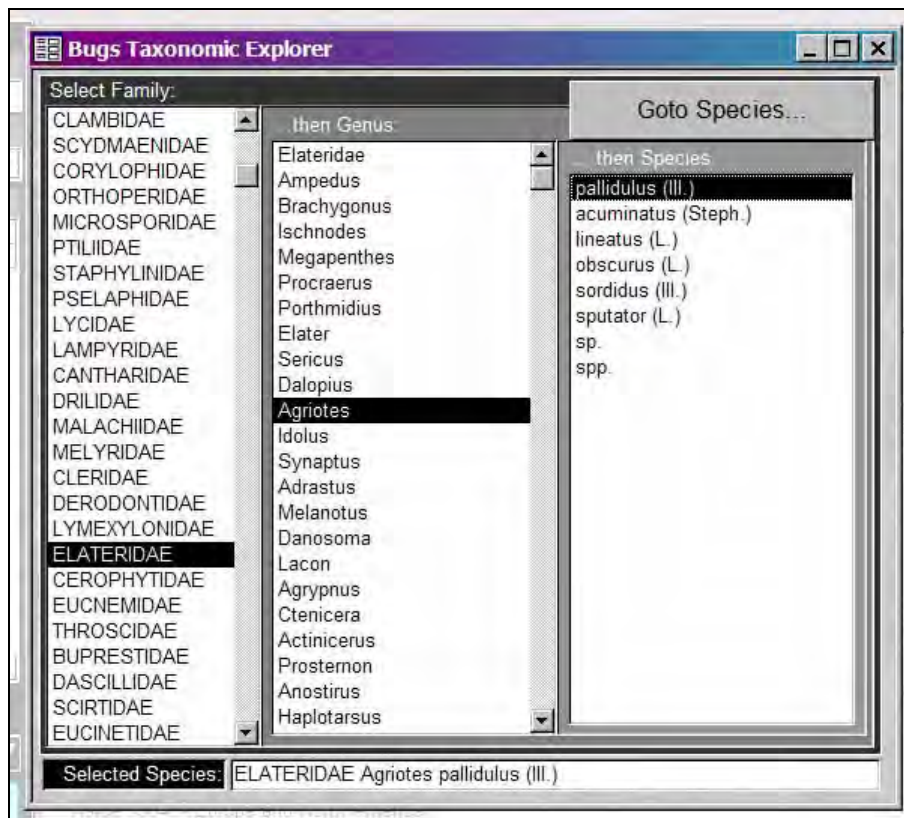


Figure 3.14. Screenshot: Taxonomic Explorer (Main Screen Navigation).

^{xix} Navigation with the keyboard requires that the containing interface, and not a subform or data box, has the focus. Using the keys when a text box is selected, for example, will scroll the selected text box, and not move the display to another taxon.

The **Taxonomic Code Browser** (Figure 3.15) fulfils a similar function, but is aimed more at users with knowledge of, or use for the taxonomic codes used to uniquely identify taxon in BugsCEP. It is particularly useful for quickly locating species where the code is available in an exported report. The [Find a code...] button on this pop-up opens the standard MS Access find dialog box (lower half of Figure 3.15), which can be used to search for a specific code rather than browsing using the scroll bars in the main pop-up. The example shows *A. pallidulus* (Ill.) selected. As in the Taxonomic Explorer, double clicking on the taxon, or clicking the [Goto Species...] button will update the main screen to display the available data for the taxon. The pop-up and dialog can then be either closed or moved.

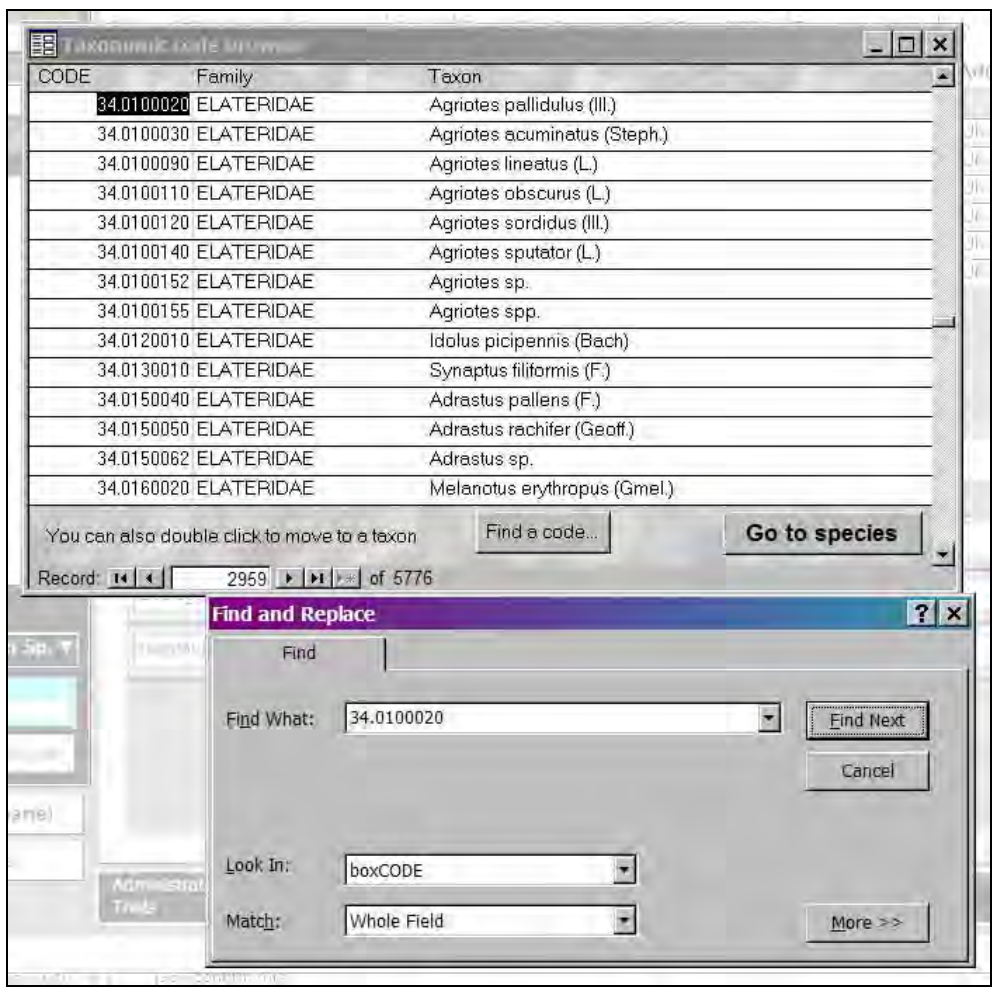


Figure 3.15. Screenshot: Taxonomic code browser (Main Screen Navigation).

3.4.1.3 Section (3): Information tabs, and section (4) Information area

Sections (3) and (4) of the main screen are intimately related in that clicking the information tabs in section (3) dictates the information displayed in section (4). Section (3) is a series of tabs which activate information panels within the information area.

The **[Synonyms] panel** is quite self explanatory (Figure 3.16), and lists the known synonyms for the selected species, with references to the source for each synonym. Note that whilst BugsCEP does refer to the synonym authority^{xx} of either genus or species, it does not always provide the primary source

^{xx} Although BugsCEP is primarily a database of ecology and fossil data, and is not designed to hold complex taxonomic information.

for the subjection of the taxon and users should use *Fauna Europaea*^{xxi}, or other information sources for detailed synonymy. Clicking on the [Bibliography for panel] button, as always, will bring up full details of the references cited (see Figure 3.12). The [Notes] field is used for any important information specific to a particular synonym use, which does not really belong under the [Taxonomic Notes] data for the species in general. Note also that there remain many problems over synonymic pathways in different countries, requiring frequent updates to BugsCEP. The authors would be grateful for notice of any errors or relevant species' reviews.

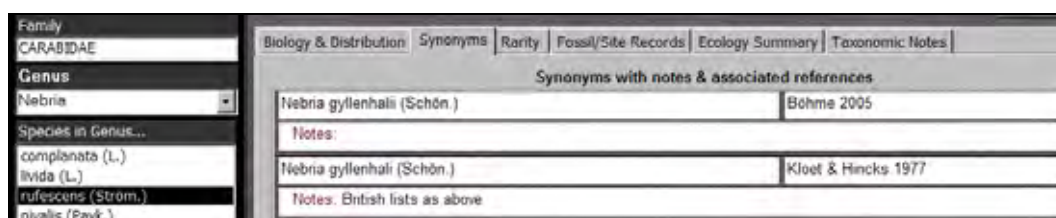


Figure 3.16. Screenshot: Main screen, synonyms panel showing data for *Nebria rufescens* (Ström.).

Synonyms are continually being revised, and the taxonomic designation used in BugsCEP has been standardized as far as possible on Böhme (2005), with a few corrections (Marshall to Marsham throughout, for example), as the most recent published European checklist; other checklists are covered in the synonyms. The use of the taxonomic code as a unique identifier means that the designations can periodically be updated without taxonomic order or links between data being lost^{xxii}. The complete list of synonyms stored for all species can also be browser using the [Synonyms Browser] button at the bottom left of section (5).

The [Rarity] panel shows the RDB status for the selected species, along with references and keys for three RDB systems, as shown in Table 3.10. The authors of BugsCEP recognize that these are not the most up to date sources or classifications, but there has not been time to update the data.

Table 3.10. Red Data Book classification systems available in BugsCEP.

System name in BugsCEP	Geographical Range	Reference
UKRDB	United Kingdom	Hyman (1992)
IUCN 2000	Sweden	Gärdenfors (2000)
IUCN 2001 3.1	International	IUCN (2001)

BugsCEP currently includes UKRDB entries for about 1 700 species. The other systems are provided for future use, and allow multiple classifications for any species so that its rarity in any number of countries can be ascertained. This could provide extremely useful in studies analysing the changes in biodiversity caused by environmental perturbations, including climate change.

The Fossil/Site Records panel is perhaps one of the most significant improvements of BugsCEP over previous versions, and perhaps one of its most useful features for palaeoecologists. The panel provides a rapid overview of the sites from which the currently selected species is known (provided that the site is stored in the database, of course). Currently, these sites are almost entirely palaeo in nature, but could equally be modern habitat studies. The current primary function of this panel is to give palaeoecologists a summary of the fossil record of a taxon, the example shown in Figure 3.17 being part of that for the cold tolerant ground beetle *Nebria rufescens* (Ström.). As mentioned previously, samples dates are stored in three categories in BugsCEP: calendar; radiometric; period. The categories

^{xxi} <http://www.faunaeur.org/>

^{xxii} The complications and structural issues surrounding the use of a taxonomic code are discussed in Chapter 2.

are displayed as a second row of tabs towards the top of the panel, and can be clicked on to reveal the different kinds of dating evidence. By default, BugsCEP will only display dated fossil records, that is to say, only information from where the current taxon has been found in a dated sample. The records are summarized as: Site name, region, country, sample name and date; and listed alphabetically by site name. Selecting the [All Samples] option button updates the display to show all occurrences of the taxon in samples, irrespective of whether they are dated or not. This is useful for getting a broader view of the occurrence of the taxon in poorly dated or modern sites. Selecting the [Dated Samples Only] switches the display back to the default view.

There were initially plans to include age depth models in BugsCEP, to allow the interpolation and extrapolation of dates between samples in stratigraphic sequences. This is under development at the time of writing, but the structural foundations have been implemented, and users can define a set of samples as a ‘Stratigraphic sequence’ when entering data (see section 3.2.2).



Figure 3.17. Screenshot: Fossil/Site Records panel, showing sites with radiometrically dated samples containing *Nebria rufescens* (Ström.).

The **Ecological Summary** panel (Figure 3.18) shows coded habitat and ecology summaries of the current species, along with information on the seasonal activity of the imagines. The two habitat/ecology classification systems, Bugs EcoCodes and Koch’s ecology codes, are summarized in Table 3.11, and described in detail in Chapter 4. The majority of palaeoentomologists have classified their species lists in terms of the modern habitats that they represent, but until now no software has been designed to apply a classification system consistently to any number of sites.

Bugs EcoCodes can be used to create a diagram summarizing the environments represented by any sample(s). This is extremely useful for summarizing environmental change over time, or comparing the signals provided by fossil insects from either different parts of an archaeological site or different horizons within a sequence. This feature is available from the BugStats program component, and is described below (3.4.3), and discussed in detail, with examples, in Chapters 4 and 6.

Seasonality data, which is also shown on this panel.

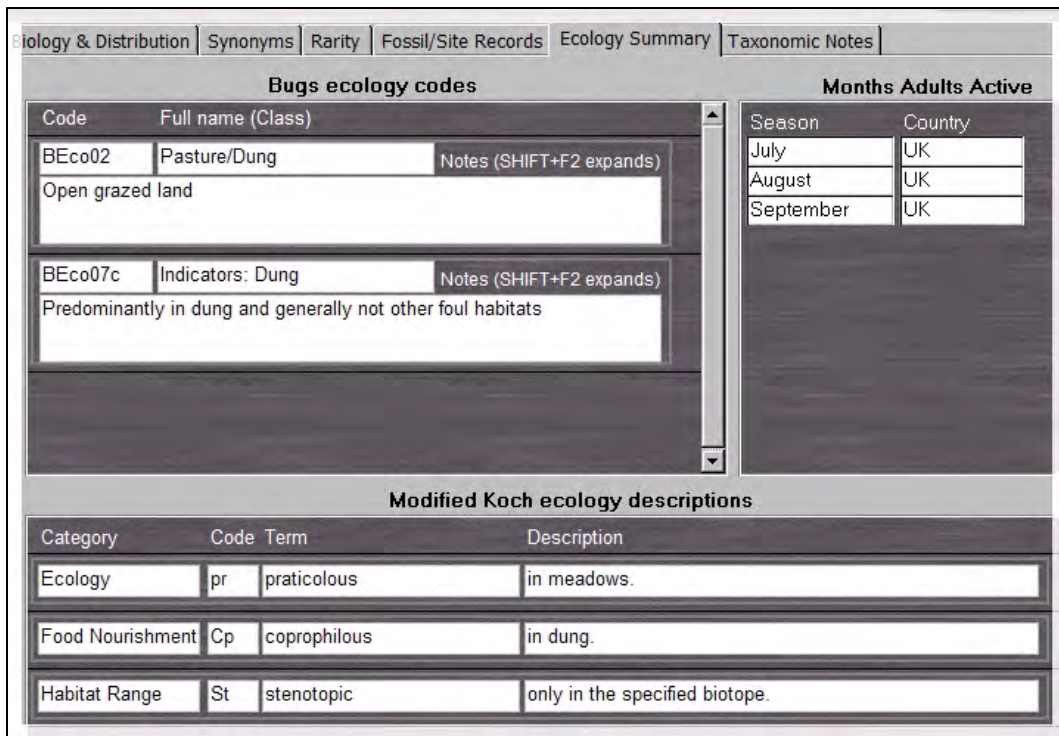


Figure 3.18. Screenshot: Ecological Summary panel for *Aphodius foetens* (F.).

Table 3.11. Basic outline of the ecology classification systems used in BugsCEP.

Ecology code system	Description
Bugs EcoCodes	Habitat summary codes, 22 classes. Devised primarily to be of use to palaeoecologists and quite generalized. Devised by Paul Buckland and Philip Buckland.
Modified Koch's ecology codes	Descriptive ecology codes, 127 classes divided into 6 categories. Give a detailed description of the ecology of the species. Devised by Koch (1989-92), and translated by Paul Buckland and Eva Panagiotakopulu. Computerized by Philip Buckland.

The **Taxonomic Notes** panel displays referenced, abstracted texts which contain information specific to the taxonomy or identification of taxa, rather than their ecology or distribution. The majority of these notes are information on the availability of identification keys, identification tips, and notes on taxonomic inconsistencies or disagreements. For example, in the taxonomic notes for the mould beetle *Lathridius anthracinus* (Mann.) Hodge & Jones (1995) note that it was “...previously confused with *Lathridius[sic] minutus*, separated on subtle structural differences and genitalia”. Clicking the [Bibliography for panel] button shows the full reference details.

The panel also shows **Measurable Attributes**, which currently includes body length data for about 2 700 taxa.

3.4.1.4 Section (5): Additional buttons and administrative controls

Section (5), in addition to holding the [Bibliography for panel] button, provides a number of buttons enabling features for both users and data managers. In the bottom right hand corner of the main screen is a small [?] button, which opens the help file for this screen (the same effect as pressing the <F1> key).

The [**Synonyms Browser**] button opens a pop-up window (Figure 3.19) which lists all the synonyms currently stored in BugsCEP. It allows a user to browse or search for a species by its synonym, and then move to the species in the main screen by clicking the [Goto Species in Main] button. Clicking the [Find synonym...] button opens a custom search dialog, with which the user can search for names, or name parts, within the synonyms list. The example in Figure 3.19 shows how ‘*gyllenhalii*’ is searched for. It should be noted that the Find Synonym dialog box may seem to disappear after clicking on one of the [Find ...] buttons, but it is simply hidden behind the Synonyms Browser as the latter is given the focus. Users can position the two pop-ups for more convenient viewing^{xxiii}.

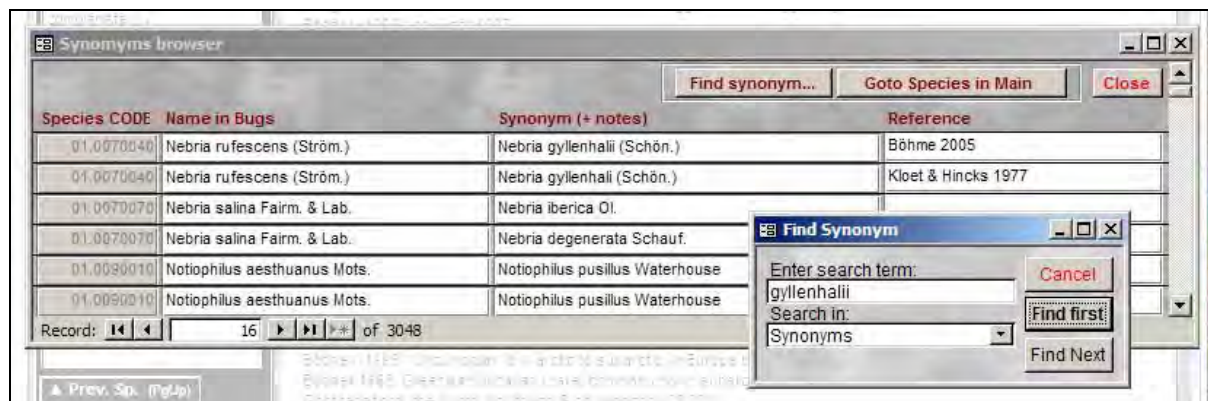


Figure 3.19. Screenshot: Synonym Browser, showing search for ‘*gyllenhalii*’. The Find dialog has been positioned to overlap the browser for ease of use.

The [Export] button on the main screen provides a variety of options for outputting the data for the current species to an external file or program. A dialog box displays the available reporting options, and the third reporting option, which exports the species name, synonyms and taxonomic notes may be of most use to taxonomists. The fourth option, which outputs the species name, biology and distribution texts along with associated references will undoubtedly be the most useful for ecologists and palaeoecologists as it provides the most descriptive habitat information, with directions to the original sources. The possibilities available on previewing the reports, along with the other report types available are described later in this chapter (3.5).

The final button visible in section (5) is the [Add/Edit Data] button, which is password protected to prevent accidental data alteration. Entering **edit mode** provides an additional [Go to CODE] button, along with buttons for adding, deleting or altering taxon designations. An [Edit Bibliography] button also appears, which is a function duplicated on the [Tools] drop down on the toolbar in section (1). Due to the large scope of BugsCEP’s data there is some variety in how different data items are added. Some items must be typed (or pasted from the clipboard), and others chosen from drop down lists. Others, such as entering references are slightly more complex and involve a number of validation checks. Currently, all data are entered, on request, by Paul Buckland on the master copy of the database, and thus the majority of users will not need to enter species related data. The issues surrounding species data entry are briefly explained in section 3.6.

Outside of the numbered screen sections (Figure 3.13), two further buttons, only one of which is visible for all taxa, are provided for accessing information on **species associations and identification keys**. The latter have been provided by Peter Skidmore (Skidmore, unpubl.), and are standard dichotomous keys. The [Keys] button is only displayed when there is a key to the current genus

^{xxiii} This is by design. Both pop-ups are set to appear in the middle of the BugsCEP window, and thus the larger one will obscure the smaller one if it is on top of it. This setting eliminates problems that could otherwise arise from different screen sizes, resolutions and numbers on different computers influencing the physical position of pop-up windows.

available, and is only active when the sp. taxon placeholders are selected^{xxiv}. The key to click beetle genus *Agriotes* (Figure 3.20), for example, is only available when *Agriotes sp.* is selected.

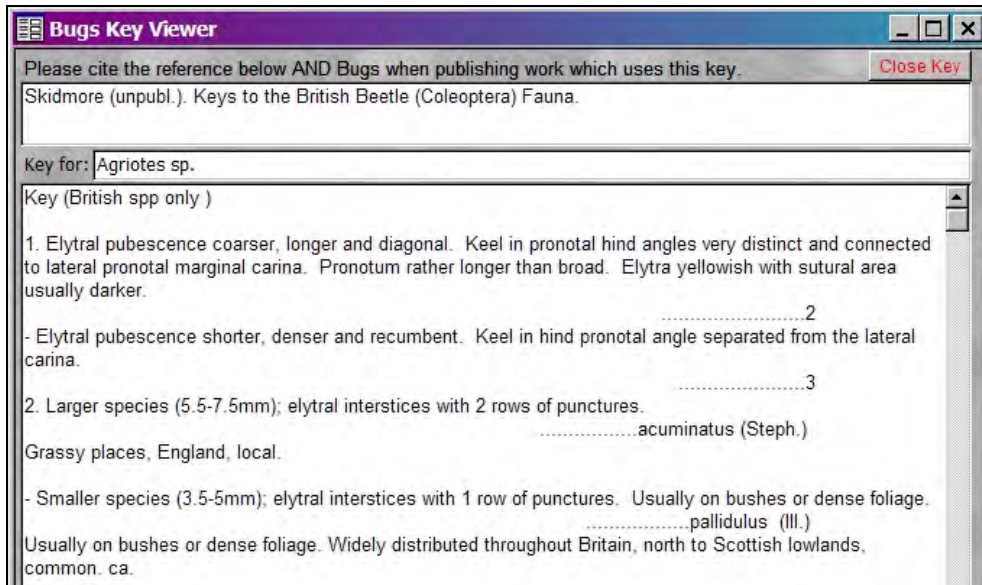


Figure 3.20. Screenshot: First two couplets of the key to *Agriotes* species.

A list of **Species Associations** is provided on clicking the [Species Associations] button, which opens a pop-up listing known relationships between the current species and others. Figure 3.21 shows the four associations stored for the halotolerant wetland/coastal species *Dyschirius nitidus* (Dej.), and clearly indicates its predatory relationship with rove beetles of the genus *Bledius*. Each and every association is referenced, and clicking the [Bibliography] button provides full details of the listed references. In this case, only Lindroth (1985) has been used to provide the raw data.

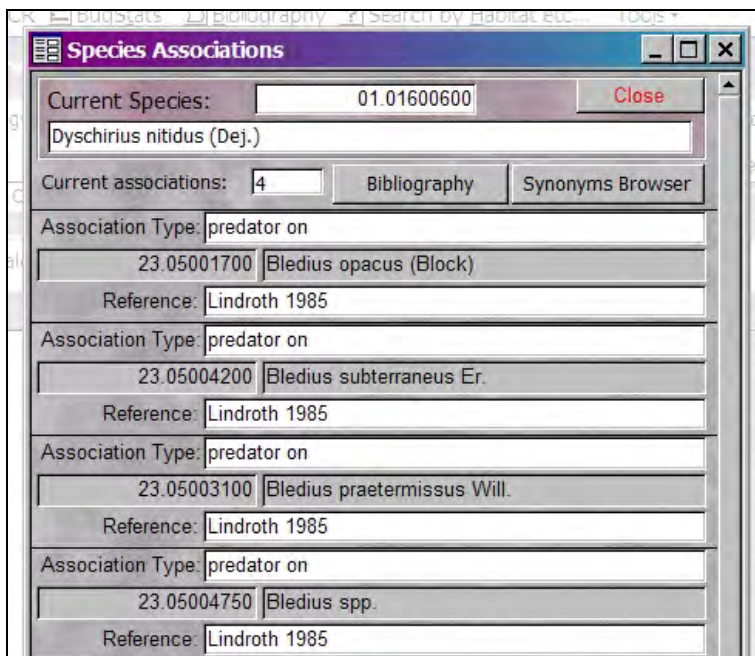


Figure 3.21. Screenshot: Species associations for *Dyschirius nitidus* (Dej.).

^{xxiv} This is not an ideal solution, and there are plans to revise the entire identification key system to include hyper-linking and images.

Clicking on the [Species Associations] button when viewing the main screen information for the first of the predated species, *B. opacus* (Block), reveals a reciprocal link to *D. nitidus* (Dej.) along with the information that it is also preyed upon by the closely related, but more eurytopic *D. politus* (Dej.) (Koch, 1989-92).

A large number of the species associations currently included are undefined, indicating non-specific associations between the species. In addition, the large number of non-coleopteran species associations is not catered for by the structure. The majority of associations included are by way of habitat, and it has not yet been decided how to standardize and store the large number of possible relationships into a manageable set. This is clearly an area for future development, and could be useful in studying the differential impacts of environmental change and introductions on species with differing positions in the same food web. For example, the Harlequin ladybird^{xxv}, *Harmonia axyridis*, introduced into the US from eastern Asia to control aphids “...in 1988, and now the most widespread ladybird on the continent”^{xxvi}, is now spreading through western Europe, and reached Britain in 2004 (Majerus *et al.*, 2006). It is now well established, out competing native ladybirds and being known to predate on a wide range of invertebrates, including other ladybirds, and is considered a serious threat to the native fauna of the UK. Monitoring of this species is underway, and BugsCEP could help by providing a system for documenting the ecology of the invasive species in its new territories.

3.4.2 Site and Countsheet Management – adding/retrieving sites and abundance data

All data manipulation relating to site data are performed through the **Site Manager** (Figure 3.22), which is available from the BugsCEP toolbar towards the top of the screen.

3.4.2.1 The Site Manager

This screen lists the name, region and country of every site stored in BugsCEP, along with a number of summary counts and buttons for accessing more specific site data. The English Bronze Age site at Brigg, Lincolnshire (Buckland, 1981), can be seen selected in the screenshot, as indicated by the highlighting in the left hand list and the site name in the ‘Currently Selected’ panel on the right. The first three summary boxes, under ‘Number of...’ enumerate the total number of countsheets, samples, abundance values (occurrences) for the site. Clicking the [Manage Countsheets & Create Reports] button opens the Countsheet Manager, which is described below in section 3.4.2.3.

^{xxv} Also called the Multicoloured Asian Lady Beetle, among others.

^{xxvi} The Harlequin Ladybird Survey: <http://www.harlequin-survey.org/>

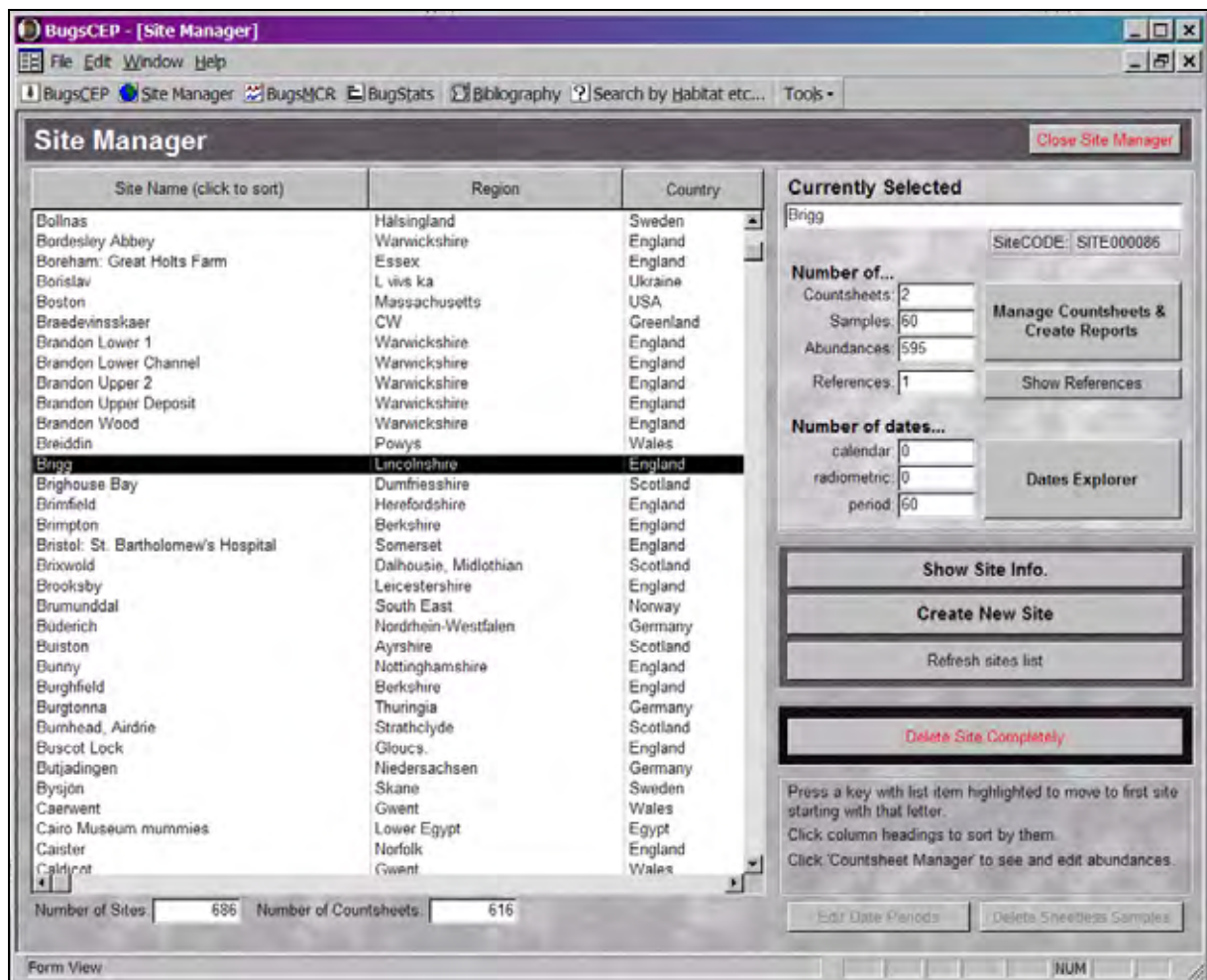


Figure 3.22. Screenshot: BugsCEP Site Manager, with the site 'Brigg' selected.

Below these summary boxes, the number of bibliographic references for the site is shown, and clicking the [Show References] button opens the standard BugsCEP Bibliography Browser with full citations. The number of dates currently assigned to the samples at the site is given in the boxes underneath the 'Number of dates' header. These are enumerated by date category, and clicking the [Dates Explorer] button will enable them to be examined in more detail (see section 3.4.2.5).

3.4.2.2 Site Information screen

Clicking on the [Show Site Info.] button opens the **Site Information** screen (Figure 3.23) with the information for the currently selected site displayed. This screen provides a basic summary of the site – its current interpretation, the names of those who did the identification work, where the specimens are stored (if known), geographical location and a simple check list of other proxy data available for the site. The display of latitude/longitude can be switched between decimal and degrees, minutes and seconds with the DD/DMS toggle button. For countries with an independent national grid, for example Ireland and the UK, a box is provided to enter this. In addition, all of the site's references are listed, as well as the countsheet records for the site, with summary data (bottom left). As before, the [Show Full References] button will open the Bibliography Browser, and the [Manage Countsheets & Create Reports] button will open the Countsheet Manager. The [Show Dates] button opens the Dates Explorer.

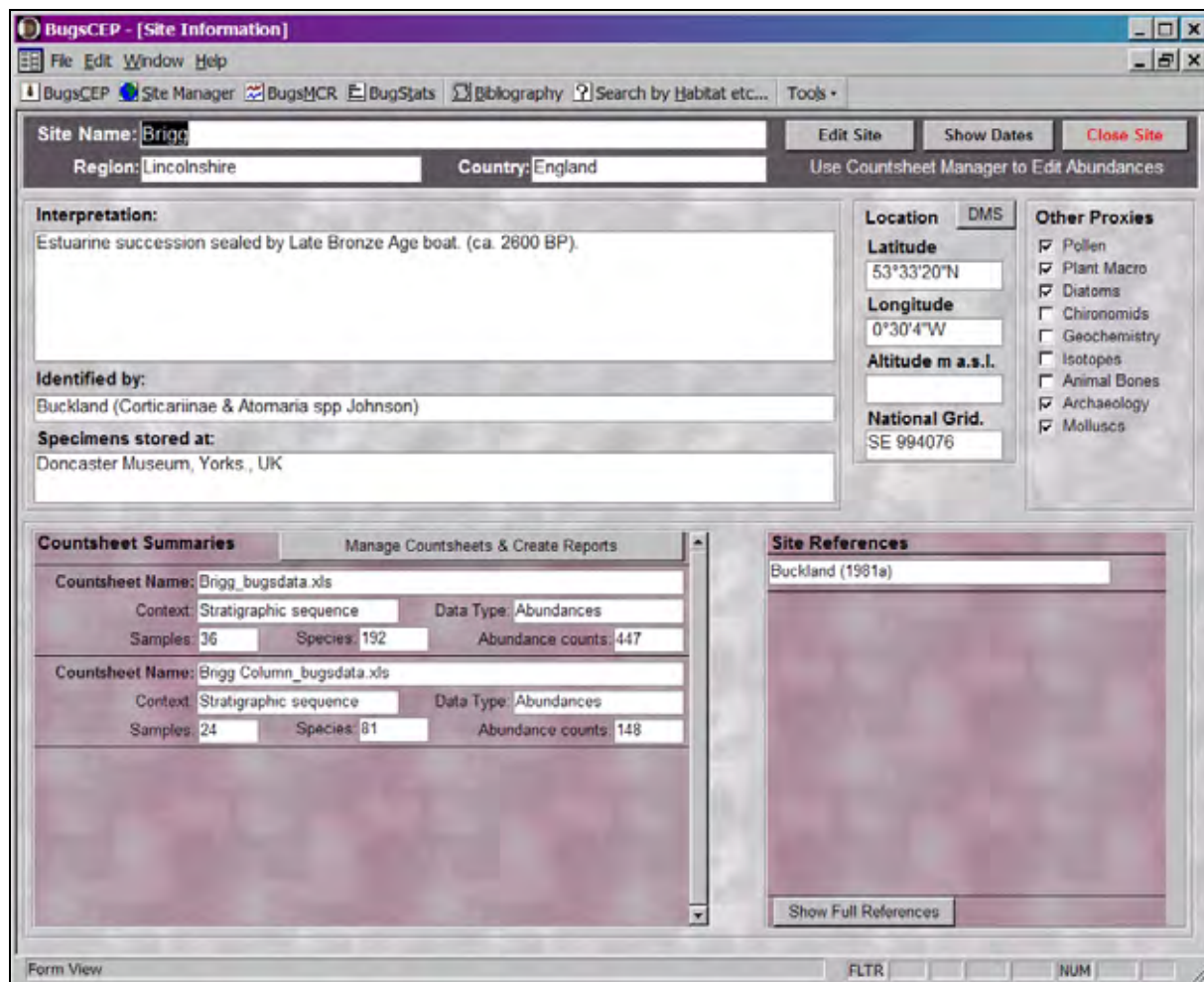


Figure 3.23. Screenshot: Site Information Screen showing details for Brigg (Buckland, 1981).

3.4.2.3 Managing countsheets and entering abundance data

The **Countsheets Manager** (Figure 3.24), accessible from the Site Manager or Site Information screens, allows sets of related samples to be compiled, stored and viewed within a site. A site may have any number of countsheets, which can be thought of as spreadsheet-like collections of species names, samples and abundance values (i.e. the abundance of species in a group of samples). If no countsheet exists for a site then the user is given the option of either creating or importing one (see next section). New countsheets must be given a name, assigned ‘context’ and ‘data type’ metadata, and then saved. Only then can samples, taxa and abundance data be added.

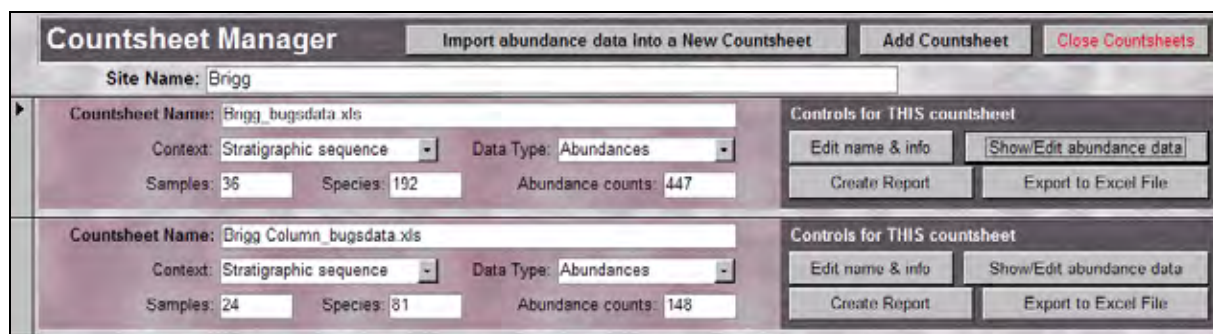


Figure 3.24. Screenshot: The Countsheets Manager showing the two countsheets stored for Brigg.

Countsheets can be examined, created, imported, exported or reported from the Countsheets Manager, and metadata describing the context and type of data stored can be edited. The Countsheet Name identifies a countsheet, although there is no demand on uniqueness for these labels. They can be named as seen fit by the user, and should be chosen so as to help others identify the data stored within them. The majority of countsheets included in BugsCEP on installation are named after the MS Excel files in which they were stored in earlier versions of Bugs. For each countsheet, three summary counts are provided: number of samples, species and abundance counts. Countsheet data can be exported as an MS Excel file by pressing the [Export to Excel File] button, but perhaps the most powerful function is accessed by pressing the [Create Report] button. Site reporting allows the species biology and distribution data, along with site summary and all relevant references, for all the species represented in a countsheet to be printed or exported. The available reports are described in section 3.5.3.

The **Countsheets Explorer** (Figure 3.25) is activated by clicking the [Show/Edit abundance data] button on the countsheet record. This is a spreadsheet-like editing area for viewing or changing species abundance data, where sample names form the column headers, apart from the first column which contains taxa names^{xxvii}. Numerical abundance data (or a '1' indicating presence only) can be entered directly into the data cells.

SpeciesName	S0-5cm	S5-10cm	S10-15cm	S15-20cm	S20-25cm	S25-30cm
Bembidion lampros (Hbst.)	1					
Agonum sp.	2					
Halipilus sp.	1					
Hygrotus inaequalis (F.)	3					
Hydroporus erythrocephalus (L.)	3					
Agabus sturmi (Gyll.)	1					
Gyrinus sp.	1					
Hydraena testacea Curtis			1			
Hydraena spp.			1			
Ochthebius dilatatus Steph.			2			
Ochthebius minimus (F.)			2			
Ochthebius spp.				1		
Helophorus porculus Bedel	1					
Helophorus grandis Ill.	1					
Helophorus brevipalpis Bedel	10					
Helophorus spp.	14		1			
Cercyon spp.	1					
Megasternum obscurum (Marsham)	1				2	1
Hydrobius fuscipes (L.)	1					
Laccobius bipunctatus (F.)	1		1			
Leiodes sp.						
Corylophus crassidoides (Marsham)						
Ptenidium nitidum (Heer)						
Acrotichis sp.						
Metopsia clypeata (Müll.)						

Figure 3.25. Screenshot: The Countsheets Explorer spreadsheet display, showing part of the abundance data for the countsheet 'Brigg Column_bugsdata.xls'. Taxa names fill the first column and sample names form the remaining column headers.

Taxa are added to or deleted from the list by clicking the [Manage Species List] button, which opens the **Species List Editor**, shown in Figure 3.26. This form consists of a taxonomic selection system

^{xxvii} Note that the standard BugsCEP menu and toolbar are not available during a countsheet editing session.

identical to that on the navigation panel of the main screen, which can be used to locate the required taxon. Double clicking the taxon name, or pressing the [Add to Species List] inserts the selected taxon into the current species list. Species are inserted in their correct taxonomic order, and attempts to inset duplicates are ignored. A previously entered taxon can be deleted by selecting it in the current list, and clicking the [Delete Selected] button – but note that all abundance data for that taxon will also be deleted. Alternatively, a taxon in the current list can be replaced by selecting it, selecting the new replacement taxon in the navigation box, and then clicking the [Replace Selected] button. For example, in Figure 3.26 clicking [Replace Selected] would replace *Agonum* sp. with *Agonum thoreyi* Dej. This is essential where re-examination of a specimen has lead to increased taxonomic precision. Clicking [Apply Changes] saves any alterations, closes the editor, and displays an updated species list in the Countsheet Explorer.

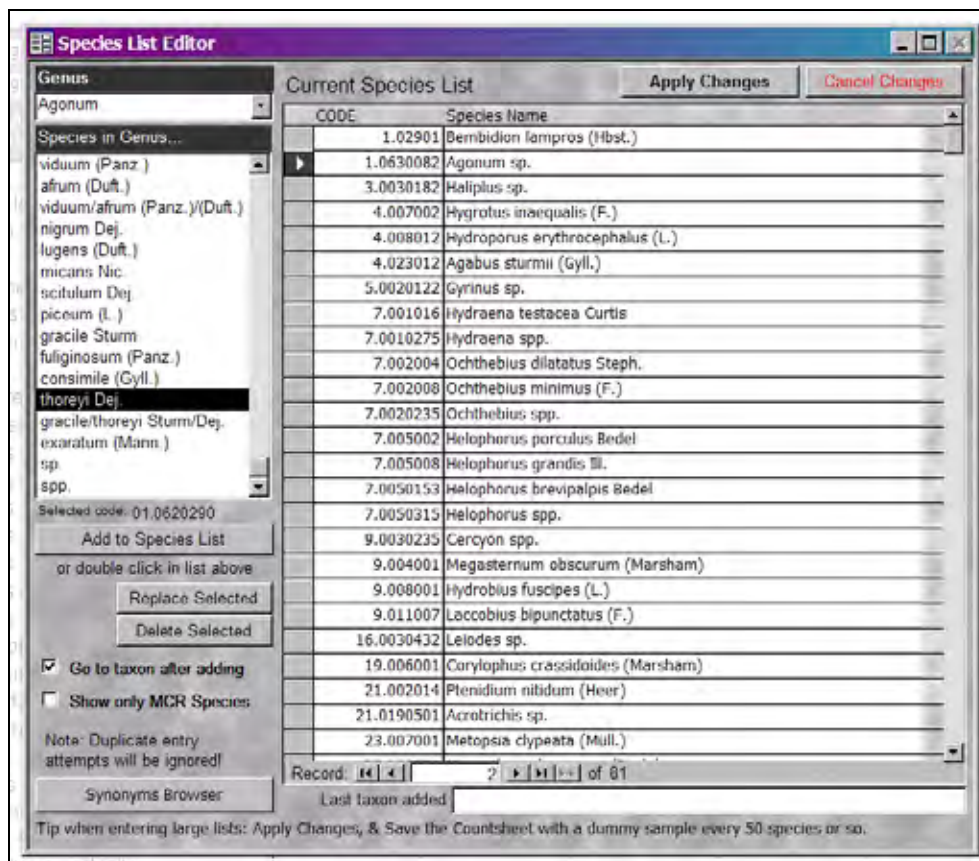


Figure 3.26. Screenshot: Species List Editor.

Samples may be viewed, added, deleted, renamed or given metadata by clicking the [Manage Samples] button and opening the **Samples Manager** (Figure 3.27). Sample depths and coordinates may be assigned to samples, and will be available in future versions for sorting samples in outputs.

The Countsheet Explorer includes a number of validation features, to prevent the input of non-numerical data, and the storage of empty species or sample rows, warning users when such attempts are made. It also includes the facility to create a ‘dummy’ sample with an abundance=1 for every species that has been entered. This feature, activated by clicking the [Create ‘Presence’ dummy sample] button, is useful when creating presence only species lists, eliminating the need to laboriously enter ‘1’ in every cell. It is also useful when a long species list has been entered, but abundance data entry must be temporarily aborted. The Countsheet Manager will not allow a countsheet to be saved if there are species without any abundance values, and the creation of a dummy sample, which can be deleted when the full dataset has been entered, allows the incomplete countsheet to be saved. The

[Export] button allows the user to export the current countsheet as an MS Excel file, for use in other software or import into another copy of BugsCEP.

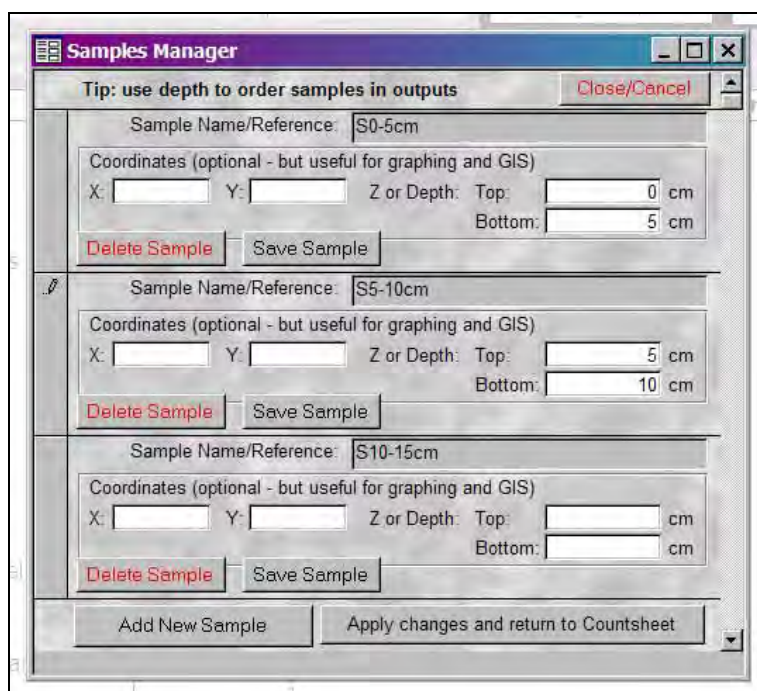


Figure 3.27. Screenshot: Samples Manager showing the first three samples of the stratigraphic sequence from the countsheet ‘Brigg Column_bugsdata.xls’. Sample depths have been entered for the first two samples.

3.4.2.4 Importing MS Excel spreadsheets

Countsheets created in other software may be imported into BugsCEP under the following conditions^{xxviii}:

1. The data are in an MS Excel file.
2. The data are in the BugsCEP format, or an importable format where the first column (A) contains Genus and species (and optionally authority), the first row (1) contains sample names, and the data cells only contain integer numbers (as in Figure 3.28).
3. A site has been created to store the countsheet.

Once a site has been created, click the [Manage Countsheets & Create Reports] button. As there is no data currently stored for the site the user will be given the option of creating an empty countsheet or importing data from an MS Excel file. If the user chooses to import a countsheet, then the **BugsCEP Import Wizard** will open, through which seven steps guide the import of external data. The seven steps, each of which incorporate validation checks on the data, must be followed in order, and give the user the chance to add additional metadata describing the samples and countsheet. Step one involves clicking the [Select countsheet for import] button, and then locating the file on the computer. If the data file is in the BugsCEP format, i.e. was created by BugsCEP, then import is quite straight forward, and Figure 3.29 shows the successful completion of the first two steps of selecting the file for import, and extracting the sample names. Step three only becomes available at this point.

^{xxviii} There are also a number of additional important considerations in terms of the format of the MS Excel file, which are explained on the website http://www.bugscep.com/help/help_enteringdata.html#importnotes and on the equivalent page of the built in help files.

Step three opens the ‘Define Samples...’ dialog box in which users may add extra sample metadata, such as depth (top and bottom) and coordinates. The sample names cannot be changed during import, as this would make it impossible to match them with the source data columns.

	A	B	C	D
1	gs	F2	F4	F6
2	Dyschirius_globosus (Hbst.)	1	0	2
3	Bembidion_obliquum Strm.	0	0	1
4	Bembidion_assimile Gyll.	0	0	3
5	Bembidion_spp.	3	1	2
6	Pterostichus_strenuus (Panz.)	0	0	1
7	Pterostichus_niger (Schall.)	1	0	0
8	Demetrias_imperialis (Germ.)	0	0	1
9	Haliplus_spp.	0	0	2
10	Agabus_spp.	1	1	1

Figure 3.28. Spreadsheet format for import of MS Excel files into BugsCEP. Note that the divider between Genus and species could also be a space.

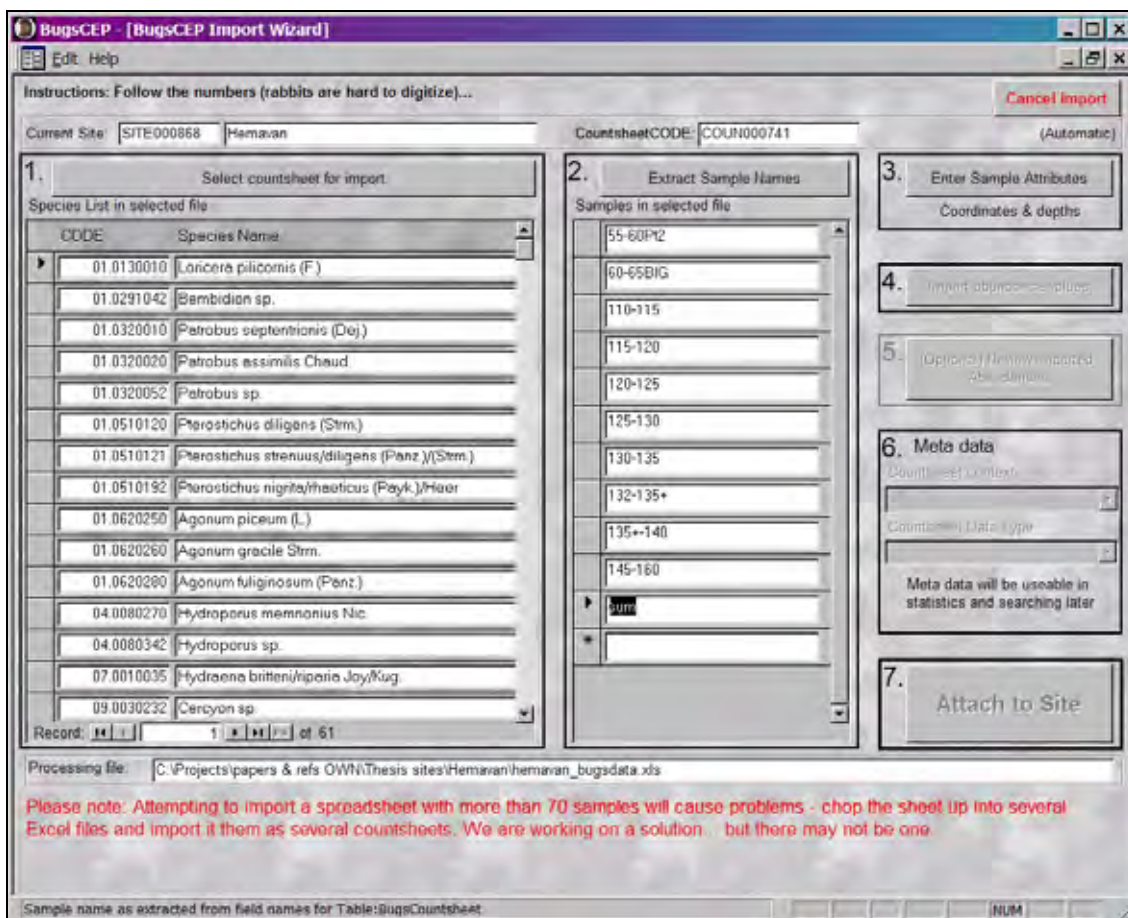


Figure 3.29. Screenshot: BugsCEP Import Wizard showing stages one and two completed in the import of the hemavan_bugsdata.xls MS Excel file.

Stage four imports the abundance data from the MS Excel file, and stage five allows the user to review these data. Stage six allows the countsheet to be assigned context related metadata. Stage seven consists of simply pressing the [Attach to Site] button, after which a ‘completed’ message is displayed, and the imported data are visible as a countsheet in the Countsheets Manager. From here it may be edited as any other countsheet.

If, during stage one, the data file is discovered not to be in the BugsCEP format, then the user will be informed and asked whether they wish to attempt to convert it. If [Yes] is chosen, then the **BugsCEP File Converter** will run, and attempt to match the species names in the import file with those stored in BugsCEP’s master index (Figure 3.30). It does this by first trying to find the genus component of the text in each cell in the first column. When it has identified a known genus, it then moves on to the species component and tries to match that. If a match is made then the taxon is stored in the New Countsheets, otherwise the cell is left blank. If a match is made, but the full name, with authority, is not the same length in the import file and BugsCEP, then it is marked with an asterisk in the New Countsheets for correction.

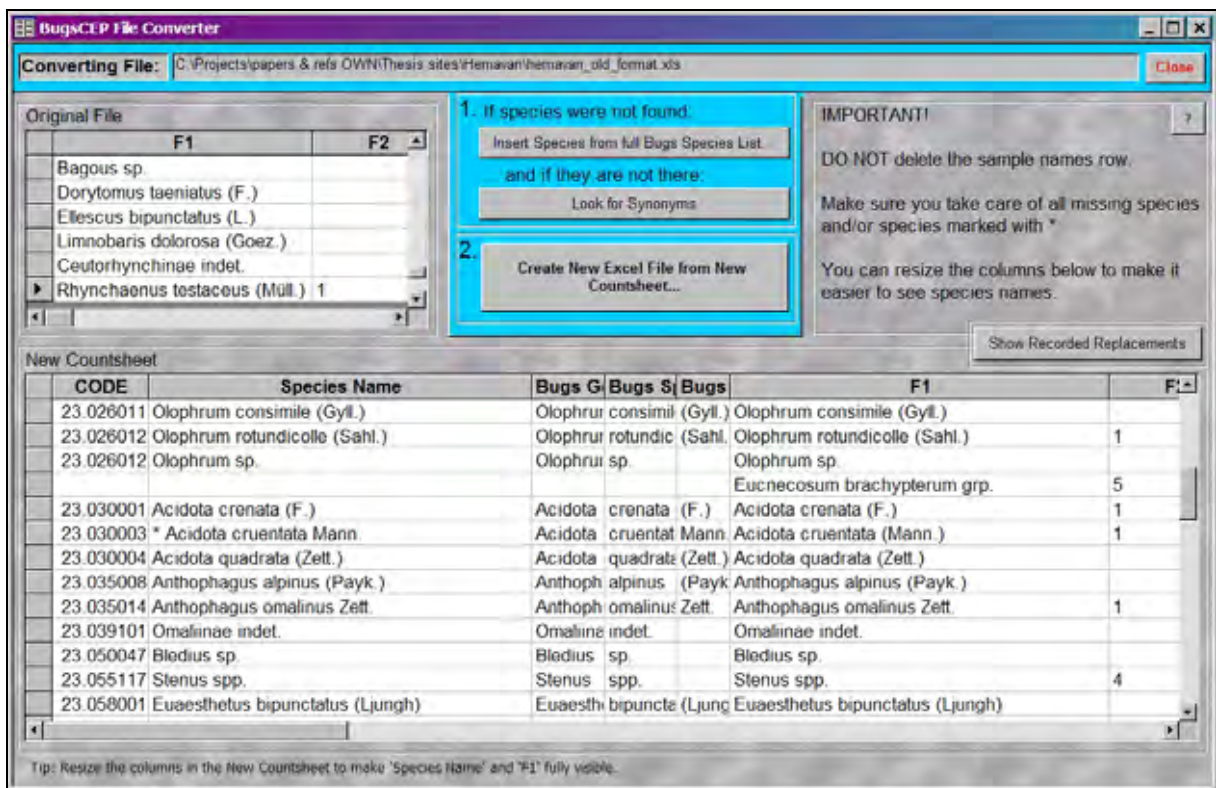


Figure 3.30. Screenshot: BugsCEP File Converter, having attempted to automatically convert the file hemavan_old_format.xls. Blank cells in the first few columns of the New Countsheets (bottom) indicate a failed match – the original name being shown in column ‘F1’. Names marked with an asterisk are uncertain matches. Column widths have been adjusted for clarity, by drag the right hand edge of the column headers.

When the automatic conversion routine is complete, a message is displayed explaining the degree of its success. In the example here (Figure 3.31), eight taxa could not be matched, and three were matched, but the converter is unsure due to differences in length.



Figure 3.31. Screenshot: Message displayed on completion of the automatic phase of imported file conversion. All problems indicated here must be corrected manually in the File Converter before import can continue.

The ‘New Countsheet’ (bottom of Figure 3.30) is used to check asterisked names, and to insert the correct taxa where no match could be made. The latter is undertaken by selecting the relevant record and searching either the BugsCEP species index or the synonyms list and inserting the correct taxon. These replacements can be stored so that the File Converter knows what to use if it encounters the same problem during a later import^{xxix}. Taxa should never be typed in manually, as this risks the insertion of an incorrect taxonomic code, and the abundance data for that taxon will be lost as the file is processed. The conversion routine is not perfect, and will miss what may seem obvious on many occasions, so patient diligence is required for its use.

3.4.2.5 The Date Explorer

The **Date Explorer**, which can be accessed from both the Site Manager and Site Information screens, is divided into two areas, the samples in the current countsheet being shown in a list on the left, and the dates for the selected sample to the right. If a site has more than one countsheet, they can be switched between using the drop down [Select Countsheet] box above the list of samples. It is also possible to toggle between the display of all samples, and only samples with dates using the [Show only samples with dates/Show all samples] button below the samples list. Tab controls are used to switch between the display of the three different date types. Figure 3.32 shows the two calendar dates stored for sample S783 at the Norse Greenlandic site of Sandnes (Sadler, 1987), indicating that the fauna in the sample most likely represent an environment present between 1000 and 1350 AD, according to historical records.

^{xxix} Full instructions are to be found in the help files or on the website at http://www.bugscep.com/help/help_enteringdata.html#import

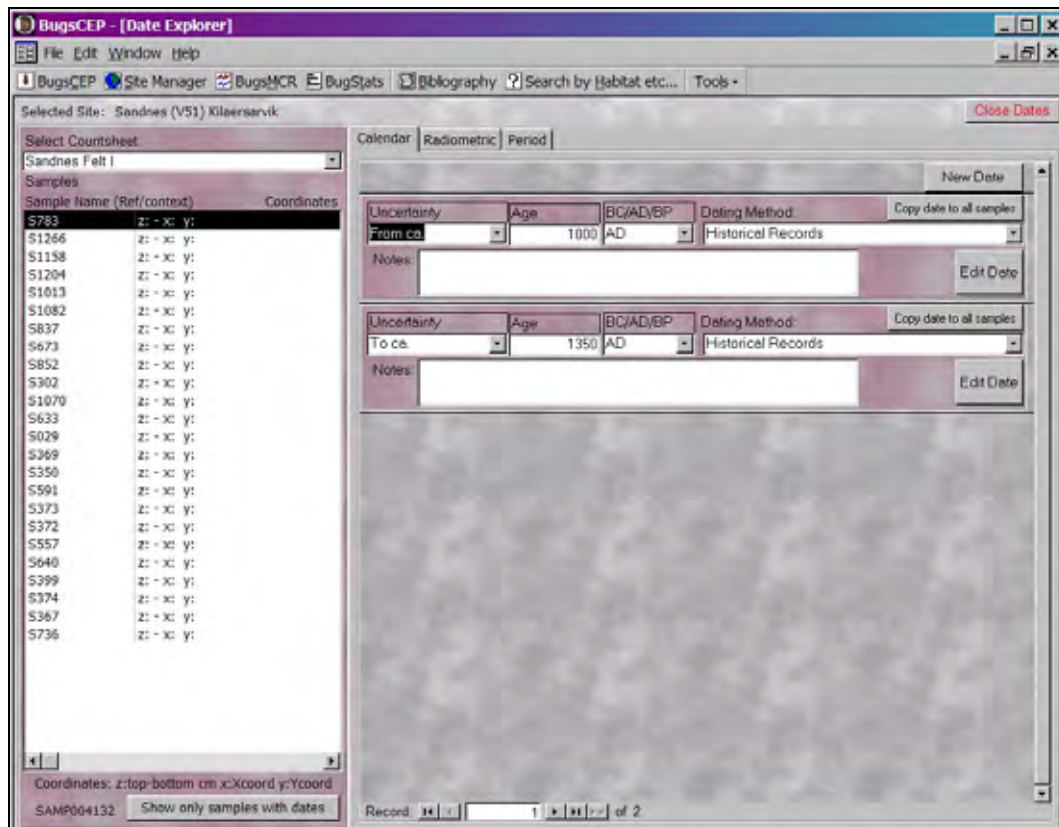


Figure 3.32. Screenshot: Date Explorer showing calendar dates for first sample in the countsheet Sandnes Felt I, Greenland.

The Lateglacial site at Messingham, Lincolnshire in the UK (Buckland, 1982), has two radiocarbon dated samples, the details of one of which are shown in Figure 3.33. A variety of dating methods are provided for, and 'other' categories are provided to allow for the storage of dates from uncommon methods. The facility to record a degree of uncertainty associated with any date is provided for all dating types, by way of drop down boxes with a limited set of choices for calendar and period dates, and as a free text box for radiometric dates. Any date may also have notes attached to it to explain further or describe the age.

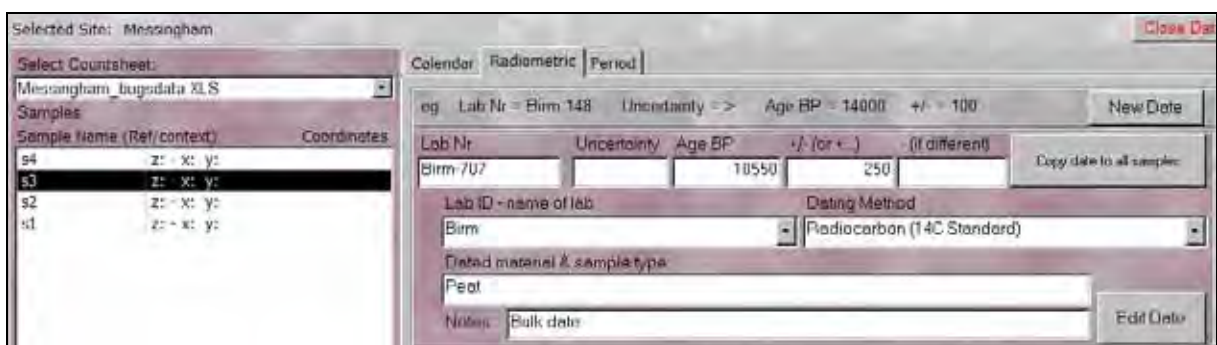


Figure 3.33. Screenshot: Date Explorer showing a radiocarbon dated sample from Messingham, UK.

3.4.3 Environmental reconstruction - BugStats

BugStats is an easy to use system for producing summaries and diagrams of the environmental implications of, or habitats represented by, the beetle taxa found at sites. It provides a number of tools and functions to assist in palaeoenvironmental reconstruction, and can help researchers to objectively compare faunas between sites and samples. This section explains the use of BugStats, whereas the

mathematics and implications of the methods used are described in detail in Chapter 4. The implications of the various calculation options are also discussed in some of the case studies in Chapter 6.

BugStats is activated by clicking the [BugStats] component button under the main menu. Sites are selected through the [Site] drop down box towards the top left of the screen. The number of countsheets associated with the selected site will then be displayed just below the name, and the [Select Countsheets] box can be used to select any of these. Note that only sites with abundance or presence/absence data are available for selection. Figure 3.34 shows the BugStats main screen with the single countsheet for the site Stóraborg selected, a complex site, which has previously been treated to a degree of numerical analysis (Perry *et al.*, 1985).

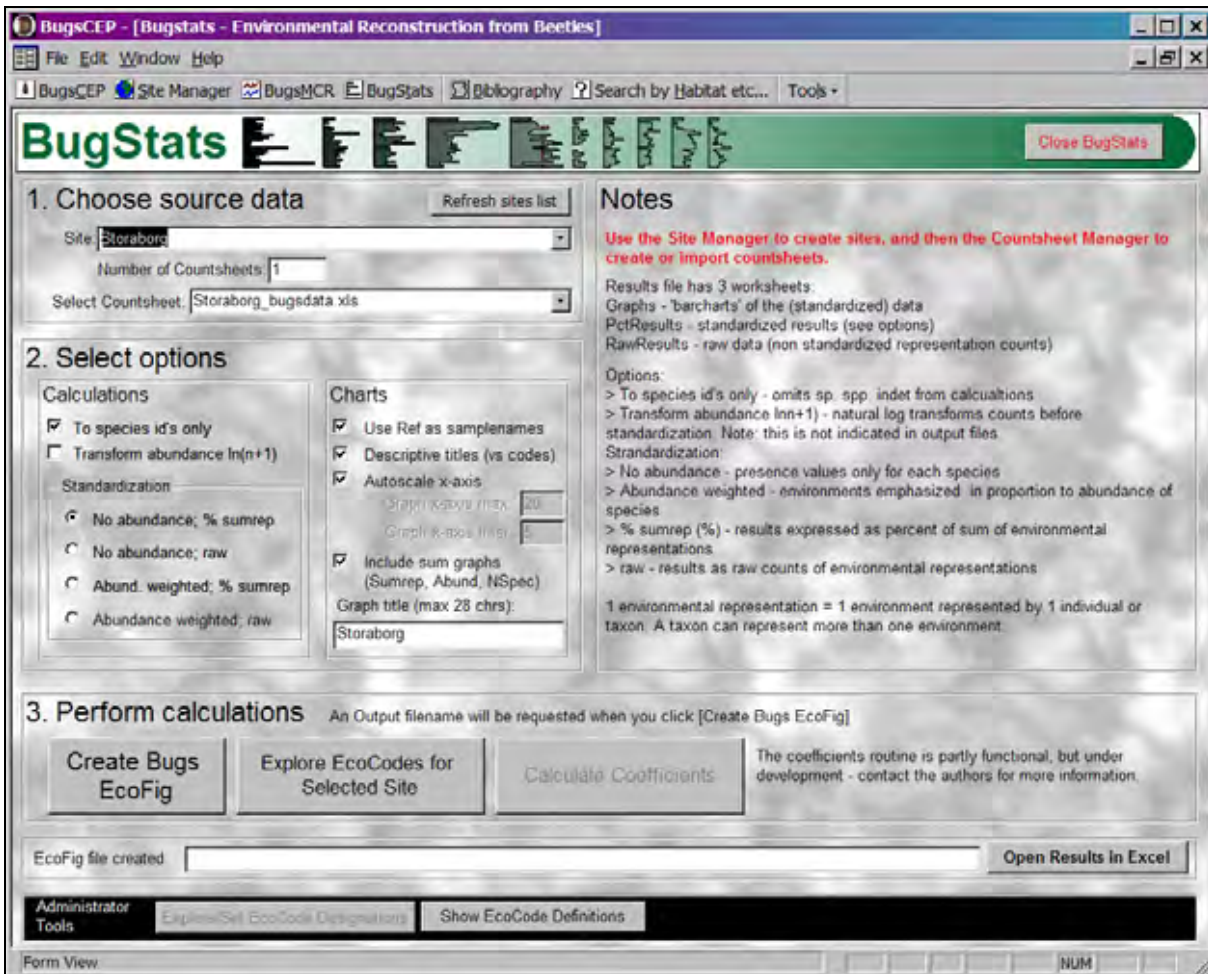


Figure 3.34. Screenshot: BugStats main screen with the site Stóraborg and countsheet 'Storaborg_bugdata.xls' selected. The 'Graph title' has been entered manually.

A number of calculation and output options are available, represented by check boxes in the 'Select options' part of the screen, and as the implications of a number of these are important for interpreting the results, they are briefly explained on the right hand half of the screen. A key concept is that of *environmental representation counts*, which are the individual indications of a habitat class given by either an individual or a taxon, depending on the options chosen. These are the raw values upon which any further calculations are based.

3.4.3.1 Calculations options

To species id's only

When this option is checked, BugStats will ignore all sp., spp. and indet. taxa, that is to say those not identified to species level. Generic identifications are common in palaeoentomology due to the variability of preservation and identifiability of beetle fragments. These taxa have been assigned to the full range of habitats represented by their component species, and as such their inclusion in calculations creates a much broader, diluted picture of the environment represented by a sample. This is not always a bad thing, and users should use their discretion and knowledge of entomology when deciding whether to use this function or not. The use of this option probably produces the most reliable output, in that only habitats specifically represented by species are displayed, whereas switching it off may increase the probability that the environments reconstructed include those that actually existed at the time of deposition. The latter is especially true for poorly resolved or small faunas, where the removal of generic taxa may result in the loss of a significant proportion of the information available in the faunas.

Transform abundance $\ln(n+1)$

This facility applies a natural logarithm transformation^{xxx} to the raw reconstruction counts before standardization (see below). Some authors believe this compensates for inter-sample variation in a number of ways, including log/normal rank abundance distributions. This may be so with certain faunas, but ideally rank abundance curves should always be produced before assuming such a translation is necessary, and the facility to do this may be built into BugsCEP at a later date. The current recommendation is *not* to transform, as this will smooth the resulting diagram, reducing peaks and raising troughs, and thus may make it difficult to observe relative changes in habitat. See Chapter 4 for more details.

Standardization

By using raw values the results reflect the absolute numbers of species and/or individuals in each sample, and samples with higher numbers of species or individuals are likely to display stronger habitat indications. Although useful, this may not be ideal where abundances vary considerably between samples, or where more than one site is to be compared. Results may be standardized so as to allow for easier inter-sample and inter-site comparisons, the method chosen here to recalculate sample values as proportions of their respective sample totals. By doing so the results will always be comparable, irrespective of site or sample. The process is explained in more detail in Chapter 4, but the four options available are outlined in Table 3.12 in the order that they are shown in the BugStats interface (see Figure 3.34).

^{xxx} The +1 prevents an error from occurring where abundance=0, and is an accepted workaround (e.g. Jongman *et al.* 1995).

Table 3.12. Outline of standardization options in BugStats.

Option	Summary
No abundance, % sumrep	Taxa occurrence only (presence/absence). Environmental representation sum for habitat class expressed as percent of sum of environmental representations for sample (sumrep).
No abundance, raw	Taxa occurrence only (presence/absence). Environmental representation counts for habitat class only.
Abund. weighted, % sumrep	Taxa environmental representation multiplied by its abundance in sample. Environmental representation sum for habitat class expressed as percent of sum of Environmental representations for sample (sumrep).
Abundance weighted, raw	Taxa environmental representation multiplied by its abundance in sample. Environmental representation counts for habitat class only.

3.4.3.2 Chart options

These settings are more or less self explanatory, and so will only be outlined here. Users should experiment in order to find the settings most appropriate for their needs.

- *Use Ref as sample names*
Use the sample Reference/Context field to label samples (recommended) rather than the internal database sample code.
- *Descriptive titles*
Use real words for habitat titles, rather than Bugs EcoCodes (see Chapter 4).
- *Autoscale x-axis*
BugStats will scale diagrams to fit all data. Uncheck this to specify a maximum and interval yourself, which can be useful if either the autoscale function proves inadequate, or when wishing to magnify the graphs.
- *Include sum graphs*
Output summary graphs showing:
Sumrep - sample environmental representation sums
Abundance - sample abundance sums (number of individuals per sample)
NSpec - sample taxa sums (number of taxa per sample)
- *Graph title*
Use this box to give the graph a meaningful title (BugStats will append the chosen standardization option to this name on output).

3.4.3.3 Creating EcoCode outputs

Two forms of output can be created for habitat reconstructions: a standard EcoFig; and a sample by sample EcoCode breakdown.

Standard EcoFig

Clicking the [Create Bugs EcoFig] button will open the standard 'Save As' dialog box, with which the output file can be given a name and saved to a particular location. After [Save] is clicked BugStats will begin creating the output file and running calculations. This will take only a few seconds on

modern computers^{xxxi}, during which time the screen will flash and the diagram will be constructed onscreen. *Do not interrupt this process in any way!* Doing so may have unexpected consequences and could lead to the computer crashing and loss of data.

BugStats will indicate when calculations are complete, and display the path and filename of the output file in the [EcoFig file created:] box at the bottom of the screen. Clicking the [Open Results in Excel] will open the file and display the first few habitat diagrams, an example is shown in Figure 3.35. The rest of the diagram can be examined by scrolling and zooming out in MS Excel, copying all the diagram components into a graphics package such as CorelDraw, or by printing it. The output file is automatically set up to print on two landscape orientated A4 pages.

As can be seen in Figure 3.35 the habitat groups (or environments) are listed along the top of the diagram, samples are listed in a column to the left, and the individual chart scales are evident across the bottom of each diagram component. The actual meaning of the bars within the charts will depend on the options used, and Chapter 4, along with the case studies in Chapter 6 should be consulted for more details.

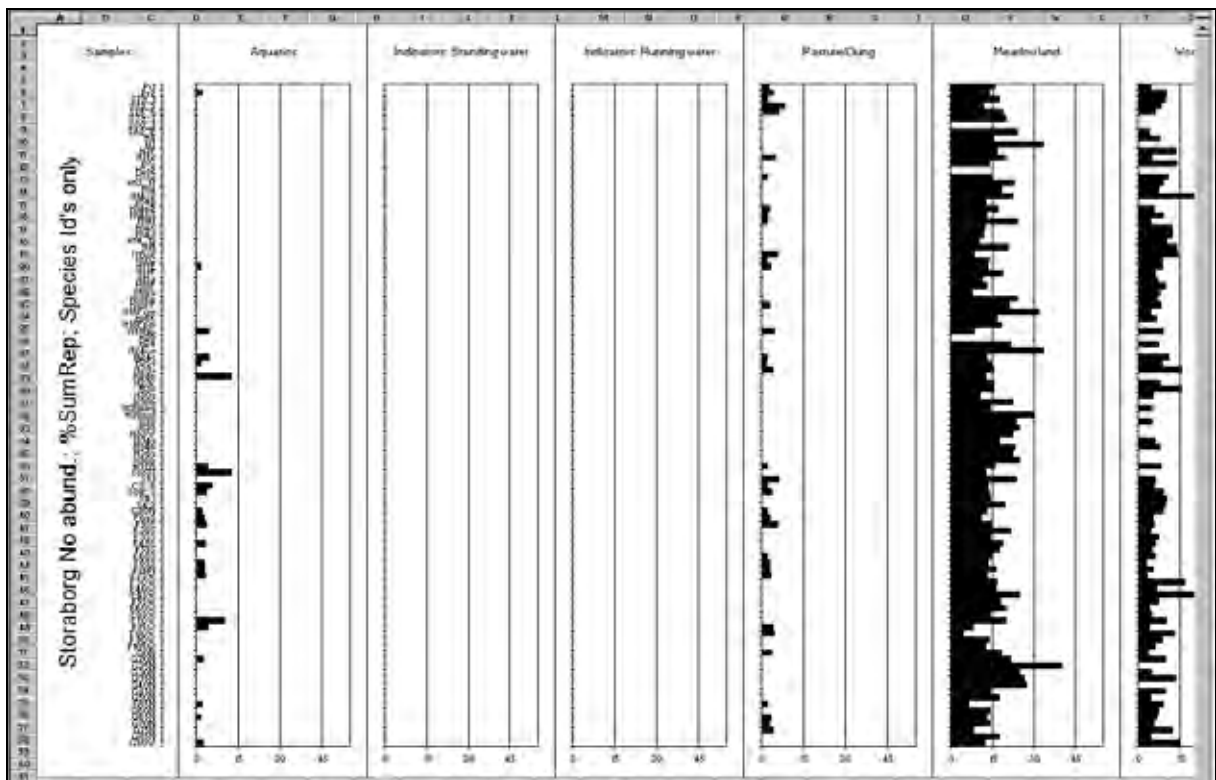


Figure 3.35. Screenshot: the first few charts of a Bugs EcoFig for Stóraborg, as opened in MS Excel for the first time.

Sample by sample EcoCode breakdown

In addition to producing the EcoFig, it may be useful to see which taxa lie behind the environmental representation counts. Clicking the [Explore EcoCodes for Selected Site] will create a report listing the taxa, their abundances and Bugs EcoCode classifications in each sample, as shown in Figure 3.36. This can help the user to understand why a particular sample indicates a particular habitat, and may also help when deciding which calculation options to choose.

^{xxxi} Faster than 2 GHz with 512 MB RAM.

This report can be printed, or exported to MS Word or Excel by clicking the appropriate button on the toolbar. The report is created as a temporary file in the currently active directory, and should be saved from MS Word or Excel to a new location.

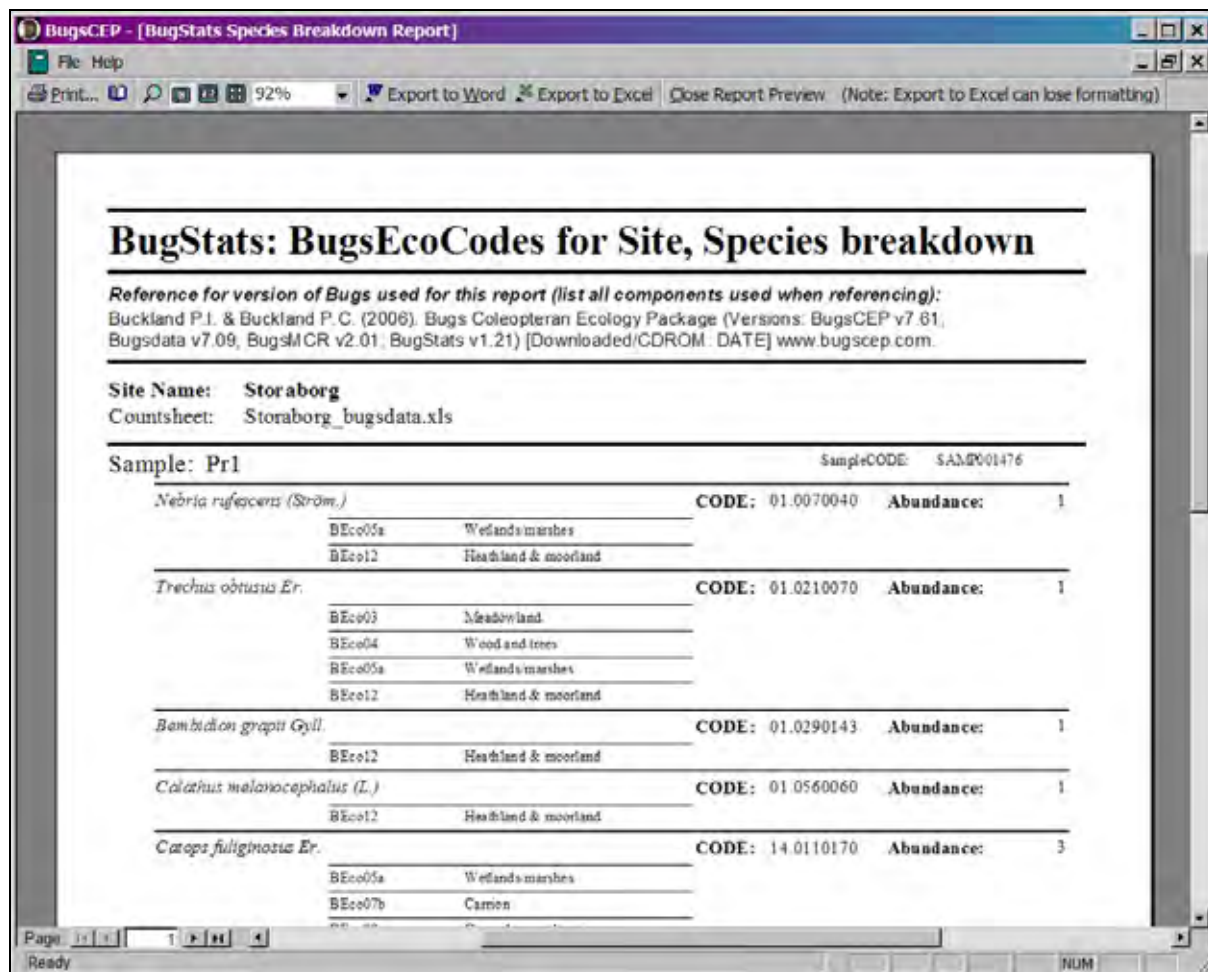


Figure 3.36. Screenshot: Start of report showing a sample by sample breakdown of the taxa found at Stóraborg, including their abundances and Bugs EcoCode designations.

3.4.3.4 Seeing the Bugs EcoCode definitions

Clicking the [Show EcoCode Definitions] button will open a form in which the formal definitions of all 22 Bugs EcoCodes can be seen. These are described in detail in Chapter 4.

3.4.3.5 Calculating coefficients of similarity

Correlation coefficients can be used to compare the similarity/dissimilarity of the species composition of samples. This is especially useful when trying to identify samples that represent similar environments in Quaternary geology and environmental archaeology, as shown in a number of the case studies in Chapter 6. BugsCEP currently only supports the modified Sørensen's coefficient of similarity (Southwood, 1978), which is the inverse (1-B) of the Bray-Curtis coefficient of dissimilarity (Krebs, 1999). The results of coefficient calculations can be used as the basis for cluster analysis in other software.

Clicking the [Calculate Coefficients] button will activate the coefficients interface (Figure 3.37), which is currently password protected and awaiting comprehensive stability testing^{xxxii}. Site and countsheet selection is by drop down boxes as usual, and the option to (natural) log transform the raw count data is available through a check box. Log transformation has a tendency to increase the observed degree of similarity when using the modified Sørensen's coefficient, which is sensitive to sample abundances (see Chapter 4 for the formula).

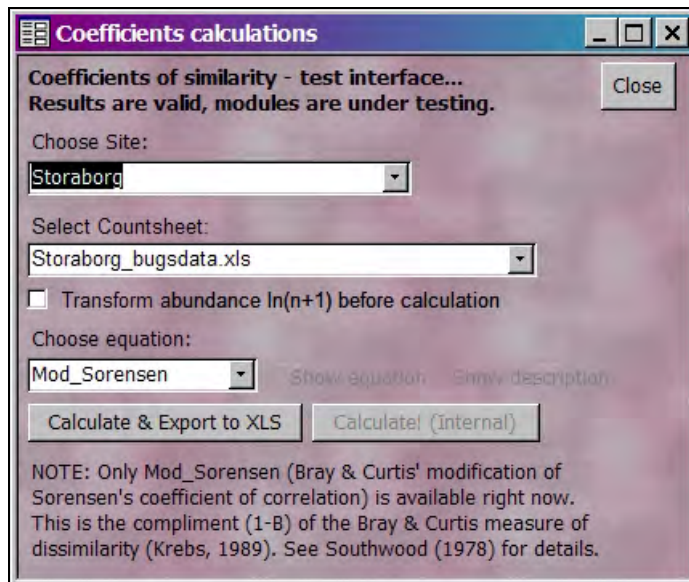


Figure 3.37. Screenshot: The coefficients calculation interface, with Stóraborg selected.

Clicking the [Calculate & Export to XLS] button runs the calculations, a 'busy' message being displayed until they are complete, after which a file name is requested through the standard 'Save as' dialog box. The screen will flash for a while as the results are exported to MS Excel, until a dialog box is displayed asking whether the user wishes to open the resultant file in MS Excel.

The results file is not formatted in any way, and users may wish to reformat the results cells as numbers with two decimal places to improve clarity. Underneath the two explanation and reference rows, the results are presented as a trellis (or matrix) diagram, sample names are listed in the first row and column, with the values in between indicating the correlation between the sample indicated by the row and column headers of each cell (Figure 3.38). Coefficient values are between zero and one, with zero indicating no similarity whatsoever between the two samples, and one indicating identical species lists in the two samples. The PopTools (Hood, 2003) 'Colour scale' function is recommended for shading the results by value.

Users should always remember that correlation is no guarantee of cause and effect, and that two samples may have similar faunas but for different reasons. It is recommended that users read standard ecological methodology texts (e.g. Southwood & Henderson, 2000; Krebs, 1999) for tips on interpreting correlations.

^{xxxii} The password may be obtained from the author, and the protection will be removed when testing is complete.

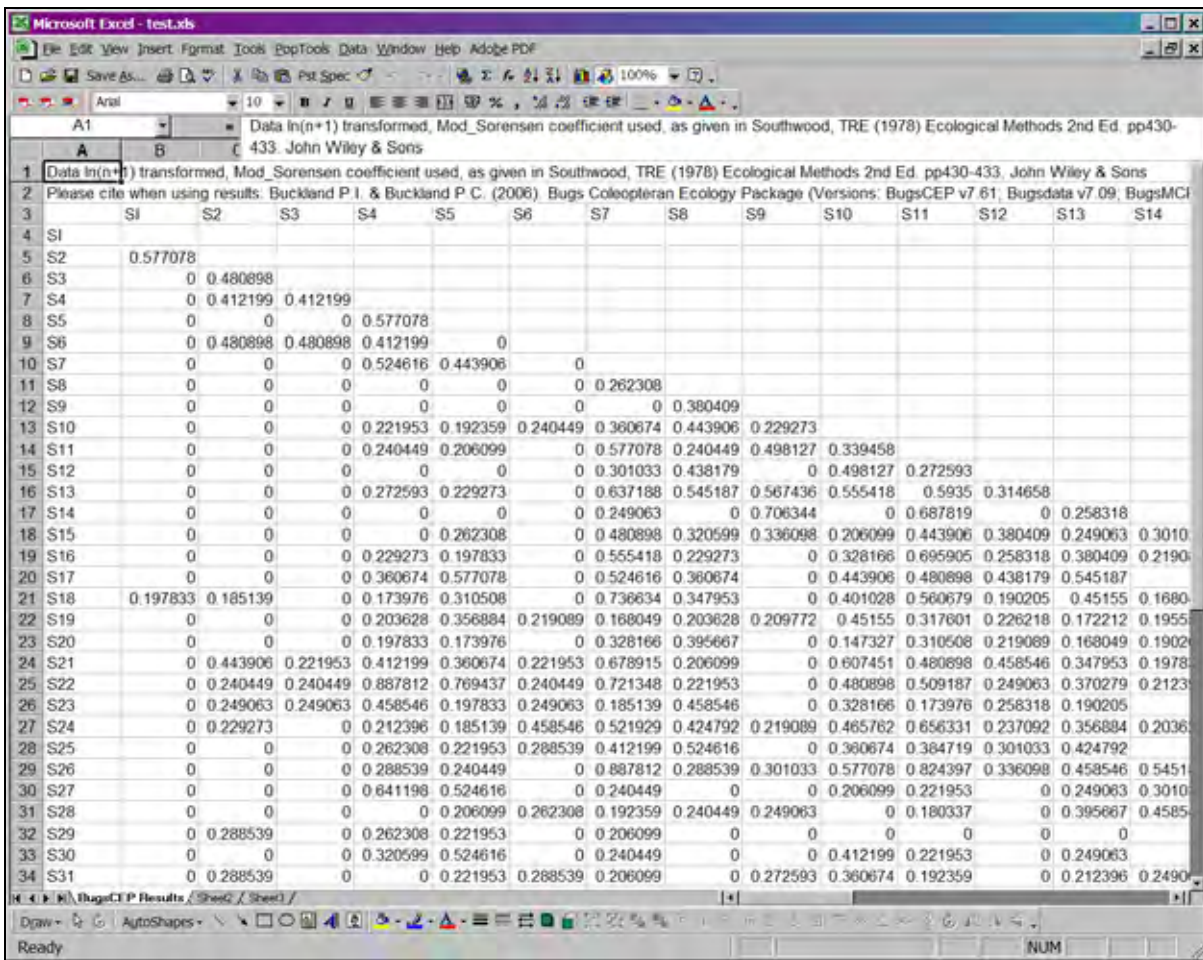


Figure 3.38. Screenshot: Coefficients output opened in MS Excel for first time.

3.4.4 Climate reconstruction - BugsMCR

The Mutual Climatic Range (MCR) method is a technique for deriving quantitative temperature reconstruction from fossil beetle remains which was developed during the 1980s (Atkinson *et al.*, 1986). It is a Modern Analogue Technique (MAT) in that it uses the modern thermal distribution of taxa as the basis for 'retrodicting' palaeotemperatures (see Birks, 1995). MCR uses the area of mutual overlap of the thermal envelopes of the species in a sample as the most probable thermal range, rather than looking for similar species assemblages from the present day, and thus circumvents the non-analogue community problems that assemblage based methods can have (Pross & Klotz, 2002). Although a number of authors have used the calibration/correction method of Atkinson *et al.* (1987), this facility is omitted here due to its unsound mathematical foundations (see Chapter 5).

BugsCEP includes the facility for running MCR calculations and outputting the results in table and graph form, along with the sample thermal envelopes (climate space maps), to MS Excel. BugsMCR uses the Birmingham 'beetle.dat' dataset (see section 2.8.5.1, and Chapter 5), and is described in detail in Chapter 5. The BugsMCR component is activated by clicking the [BugsMCR] button on the main toolbar, and all temperatures are given in degrees Celsius

BugsMCR is operated in a similar fashion to BugStats, in that the source data (site and countsheet) are selected, calculations and output options chosen, and then the calculations run. A site is selected using the drop down [Site:] box at the top left, the number of countsheets stored for the site being displayed underneath on selection (Figure 3.39). The [Select Countsheets:] drop down box can then be used to select between different countsheets if required. In a slight difference from BugStats, the selected countsheet must then be activated by pressing the [Activate this countsheet] button. This loads the full

site species list into the ‘Active Species List’, and allows the user to see how many MCR species^{xxxiii} are present by pressing the [Show only MCR species] button. In Figure 3.39 the species list for the Saint Bees Lateglacial site in Cumbria (Coope & Joachim, 1980) has been activated, and shows a total of 282 taxa. Reducing this to only MCR species reveals that only 114 of them have climate data at the current time. This is to be expected, as there are only 436 taxa in the MCR database, a full list of which, along with their RECON equivalents (see Chapter 5) can be seen by clicking the [Show All MCR Species](#) link in the Tools box. The currently active countsheet data, in their reduced MCR form if the [Show only MCR species] button has been pressed, can also be seen by clicking the [Show Current Countsheet](#) link just below this.

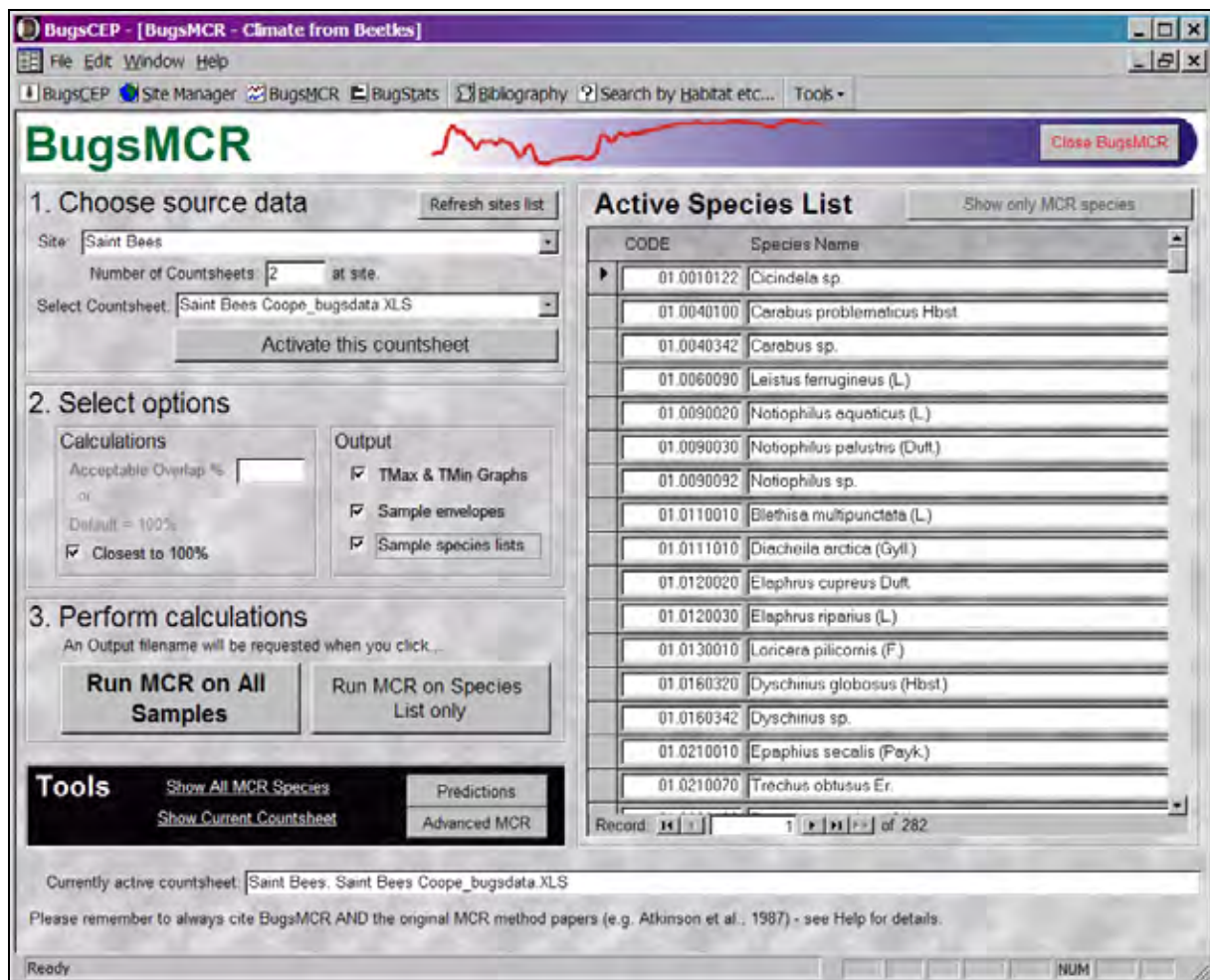


Figure 3.39. Screenshot: The BugsMCR interface with the Saint Bees site and ‘Saint Bees Coope_bugsdata.XLS’ countsheet activated.

Two calculation setting options are available in BugsMCR, the implications of which will be explained in Chapter 5, and only the usage is described here. If the [Closest to 100%] box is checked then BugsMCR will attempt to calculate a thermal reconstruction from the maximum extent of the area of highest overlap for each sample, allowing a greater scope for reconstruction than the traditional >90 % method. This does, however, come with the warning that reconstructions for the samples with less than 100 % overlaps are probably less reliable. Unchecking this box, and entering a number in the [Acceptable Overlap %:] box allows the user to specify the minimum level of overlap, in percent, on which a reconstruction should be undertaken. For the most reliable calculations this should be set to 100 %. Should the area of maximum overlap be non-contiguous, BugsMCR calculates using the

^{xxxiii} Taxa that have thermal envelopes stored for them.

extreme limits of the combined areas. Note that the default setting for BugsMCR is ‘closest to 100%’, as this allows the user to examine the results from all viable samples, and decide themselves whether to recalculate with further restrictions. The user may also delete results from the output files manually, and the graphs will be updated automatically.

The output options, to the right of the calculation options (see Figure 3.39), control the creation of two possible results files. Checking the [TMax & TMin Graphs] option, which is selected by default, graphs the results in standard MCR diagram form, although at this point in time the X, or sample axis, can unfortunately not be scaled by time or depth (Figure 3.40). Only the numerical results are exported if this box is unchecked (Table 3.13, with an explanation of headers in Table 3.14).

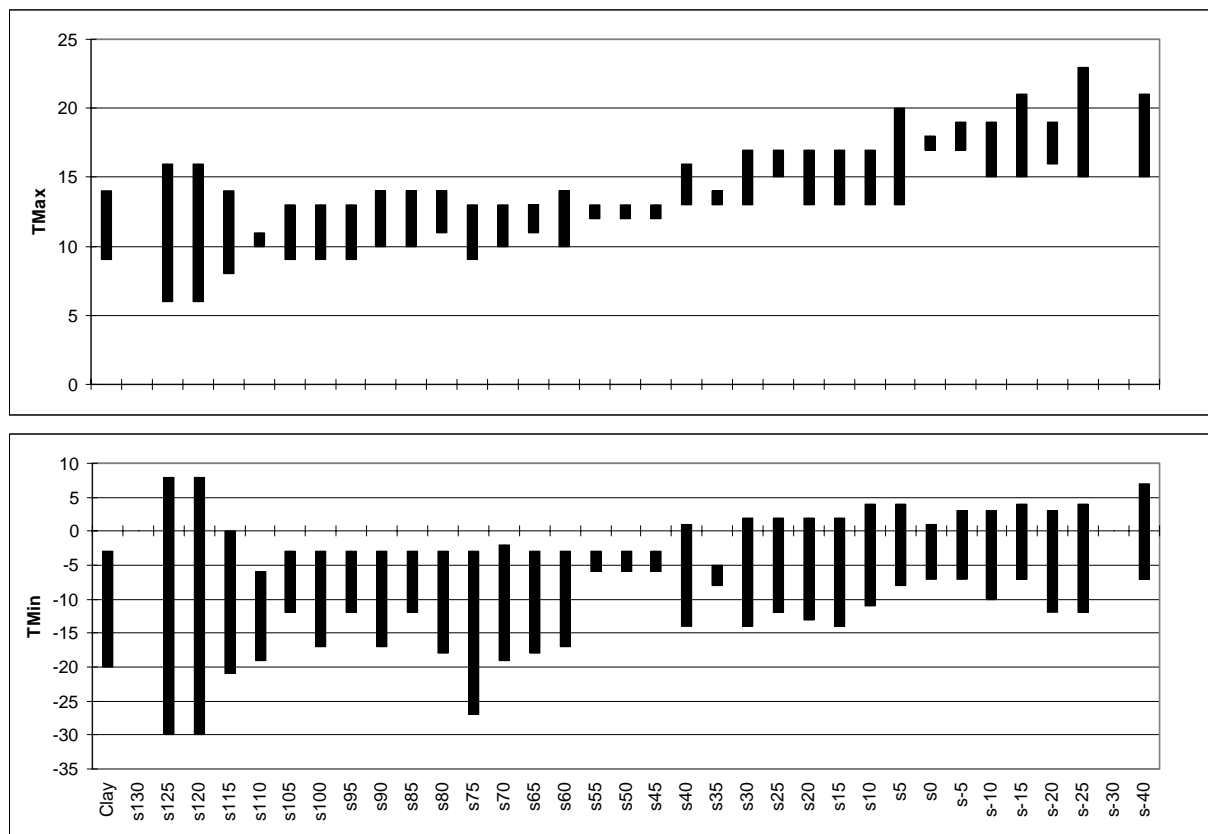


Figure 3.40. Thermal reconstruction from the Saint Bees site, using the ‘Closest to 100%’ overlap option. Note that the temperature axes may differ on output, and must be scaled manually. All results are in degrees Celcius, and the warming of both summer and winter temperatures through the sequence can clearly be seen from left to right.

Table 3.13. Extract of the numerical results output for the thermal reconstruction from the Saint Bees site, using the 'Closest to 100%' option. See Table 3.14 for an explanation of column headers.

Sample	TMaxLo	TMaxHi	TMinLo	TMinHi	TRangeLo	TRangeHi	NSPEC	Overlap
Clay	9	14	-20	-3	16	29	10	90
s130							0	
s125	6	16	-30	8	8	36	1	100
s120	6	16	-30	8	8	36	1	100
s115	8	14	-21	0	14	29	3	100
s110	10	11	-19	-6	17	29	11	100
s105	9	13	-12	-3	16	22	11	90.90909
s100	9	13	-17	-3	16	29	13	92.30769
s95	9	13	-12	-3	16	22	21	95.2381
s90	10	14	-17	-3	16	29	24	83.33334
s85	10	14	-12	-3	16	23	20	90
s80	11	14	-18	-3	16	29	18	88.88889
s75	9	13	-27	-3	16	36	15	100
s70	10	13	-19	-2	15	29	17	94.11765
s65	11	13	-18	-3	16	29	17	94.11765
s60	10	14	-17	-3	16	29	23	91.30434
s55	12	13	-6	-3	16	18	11	100

Table 3.14. Explanation of MCR results terms, see Chapter 5 for more details on the method. All results are in degree Celcius.

Column	Explanation
Sample	Sample name from database
TMaxLo	Lower limit of reconstructed mean temperature of warmest month
TMaxHi	Upper limit of reconstructed mean temperature of warmest month
TMinLo	Lower limit of the reconstructed mean temperature of the coldest month
TMinHi	Upper limit of the reconstructed mean temperature of the coldest month
TRangeLo	Lower limit of the reconstructed mean temperature ranges (TMax – TMin)
TRangeHi	Upper limit of the reconstructed mean temperature ranges (TMax – TMin)
NSPEC	Number of taxa used in reconstruction
Overlap	Percentage of sample taxa in the area of maximum overlap, used to calculate the temperature values

Checking the [Sample envelopes] box activates the creation of the second results file, which is automatically named (see below). In this file the climate space maps for each sample are exported, on individual worksheets, showing the percentage of sample species that could be found in each one degree climate cell (Figure 3.41). If the [Sample species lists] box is also checked, then the species list for each sample will be inserted under each map. These allow the user to gain a greater understanding of the thermal regime indicated by the faunas, and is particularly useful when teaching MCR theory.

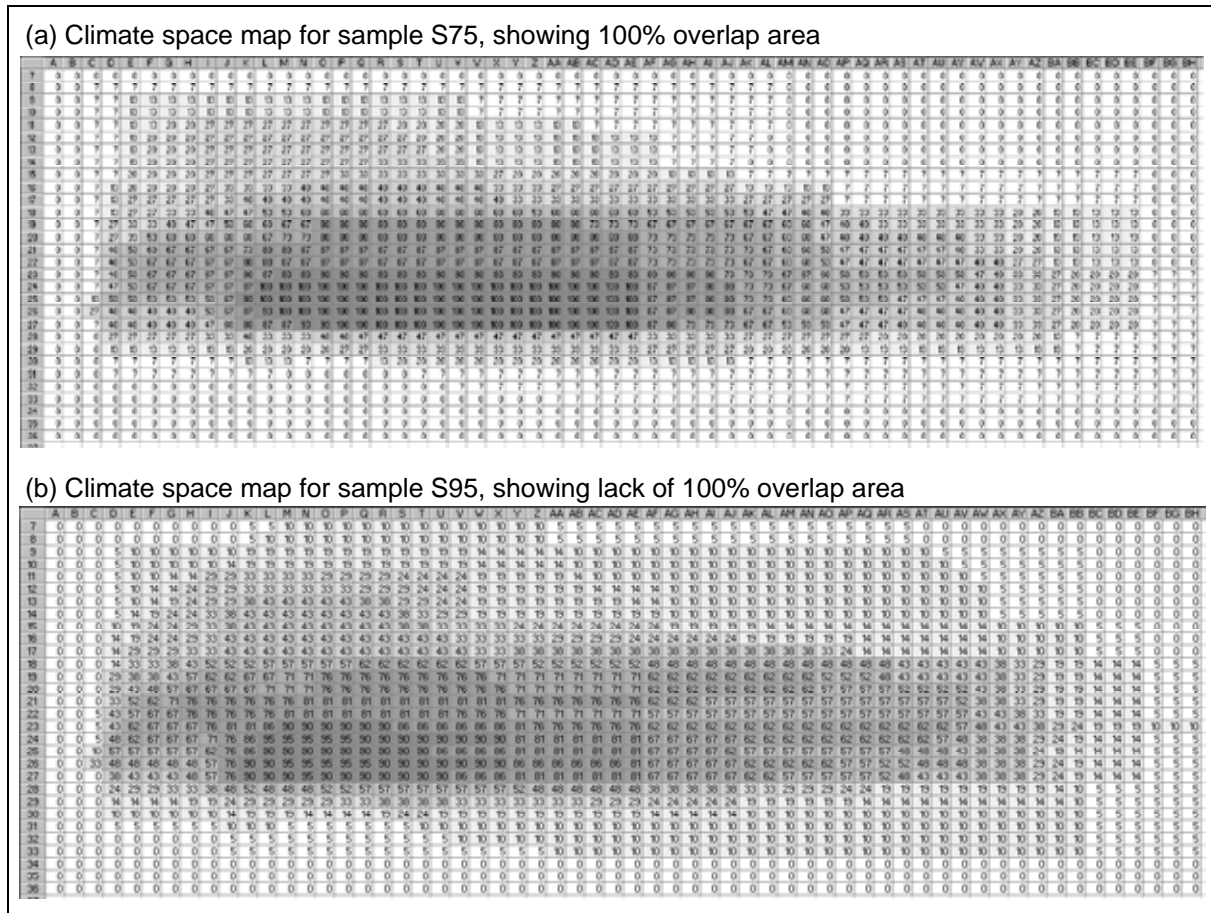


Figure 3.41. Climate space maps for two samples from Saint Bees, as exported by BugsMCR, showing (a) a 100 % overlap scenario and (b) a non 100 % overlap scenario with a complex area of maximum overlap. See Chapter 5 for further a more detailed explanation of envelopes and overlaps.

Two buttons start calculations: [Run MCR on All Samples] and [Run MCR on Species List only]. The latter of these ignores all samples and calculates temperature values for the entire species list as if it was one sample. This is primarily useful in teaching, using hypothetical data sets. The former button calculates temperature values for each sample that has at least one MCR taxon. Clicking on either button will open the standard ‘Save as’ dialog box, with which a save location and file name can be specified. A pop-up window (Figure 3.42) then indicates the progress of calculations (by way of a moving penguin), which may take some time for large datasets. If the export of sample envelopes is chosen, then this process will be repeated as the second output file is created. This second file is automatically given the same name as that specified by the user, but with the suffix “_matrices” added.

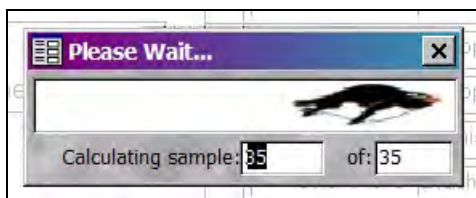


Figure 3.42. Screenshot: MCR calculation progress indicator, showing that sample 35 out of 35 is being calculated.

A message is displayed on completion of the calculations, from which the user can open the results files in MS Excel. This message must be closed in order to continue using BugsCEP.

BugsMCR also contains a number of advanced features, including species predictions and jackknifing, which are presented in Chapter 5.

3.4.5 The Search Interface

The BugsCEP search system is activated by clicking the [Search by Habitat etc...] component button on the main toolbar. Search settings are remembered between sessions, and if a previous search is found then the user will be asked whether they wish to “Continue from previous search session?” Clicking [No] leads to a request for confirmation with the question “Reset all search criteria, results and logs?”, and answering [Yes] to this opens the **Bugs Search Explorer** in its primary state (Figure 3.43).

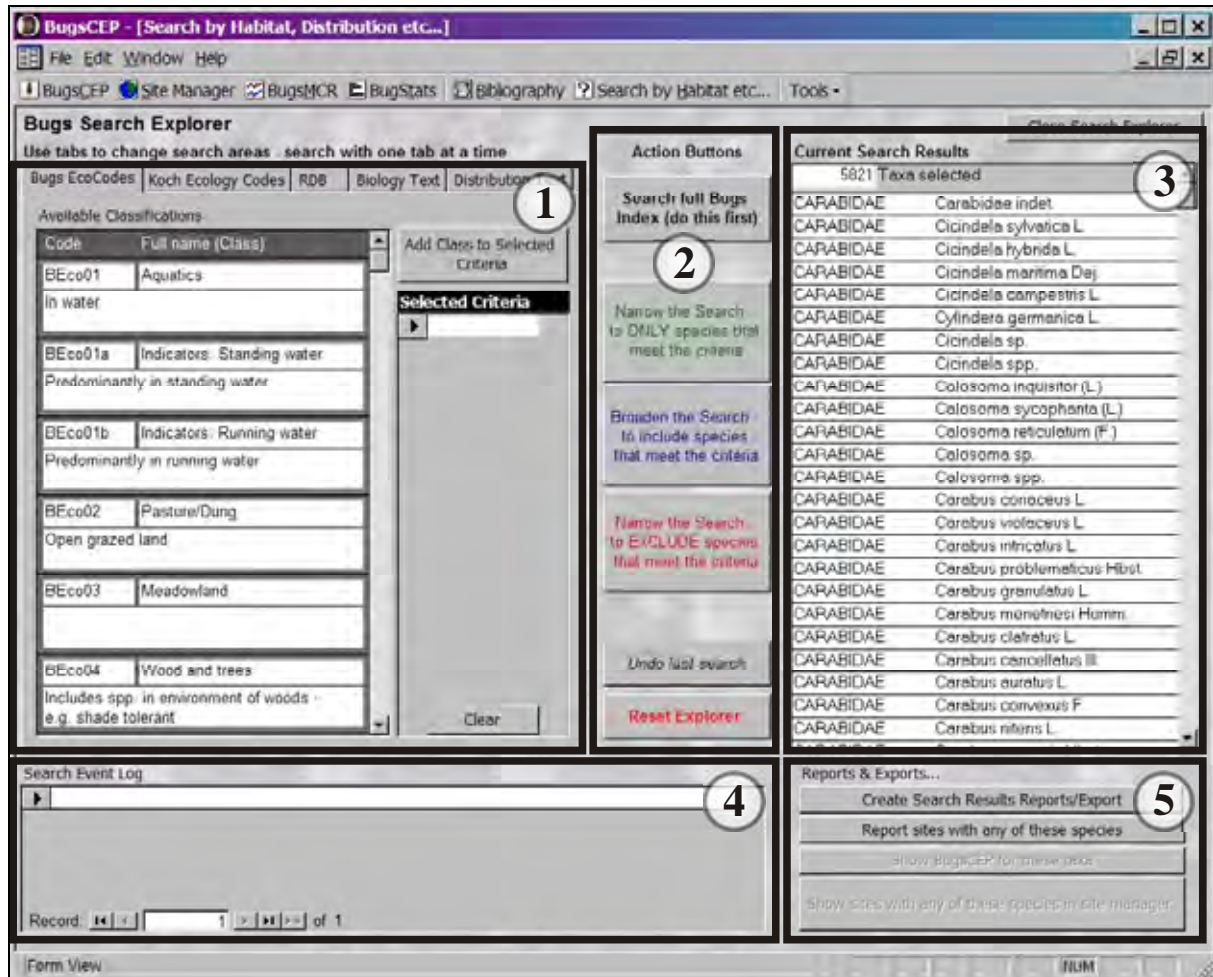


Figure 3.43. Bugs Search Explorer, with interface parts as follows: (1) Criteria type tabs (top) and search criteria selection for the active tab; (2) Action buttons; (3) Current species list (search results); (4) Current search session log; (5) Report/export buttons.

The Search Explorer is a powerful interface for retrieving species subsets through the successive application of search criteria to lists of species. The criteria type tabs, Figure 3.43(1), are used to select the area of the database to be queried, and then search criteria are either selected or entered, and then applied. The searchable data areas, along with the method by which criteria must be specified, are shown in Table 3.15.

Table 3.15. Data areas currently searchable with the Bugs Search Explorer, along with the method of criteria specification.

Search area	Method of criteria specification
Bugs EcoCodes	Code/term selection
Koch Ecology Codes	Code group and then code/term selection
RDB (Red Data Book)	RDB system and then classification selection
Biology Text	Free text entry (search terms can be saved)
Distribution Text	Free text entry (search terms can be saved)

Criteria must be selected/entered and applied using one tab at a time, although criteria are retained when the active tab is changed. On each search area tab the [Add Class to Selected Criteria] or [Add entered term to Selected Criteria] button adds the selected/entered term to a list, which is then used as the search terms when querying the current species list. Search terms are applied to the current species list, which may be the full BugsCEP species list or a result of a previous search, by clicking the action buttons, Figure 3.43(2). If multiple terms are specified, then all of them are searched for independently (i.e. logical ‘OR’ between terms). The resulting species are then shown in the ‘Current Search Results’, Figure 3.43(3), at the top of which can always be seen the number of taxa in the currently active list. Figure 3.44 illustrates the process of performing searches in the form of a flow diagram, and may help explain the logic of the process.

On each click of an action button a *Search Event Log* is created and displayed in the bottom left of the screen Figure 3.43(4), which includes the search terms used, the logic applied to the query, and the number of taxa that resulted from the search action. This log is essential for understanding how the currently displayed species list has been arrived at, and is exported with the reports available through the report/export buttons Figure 3.43(5).

It should be noted that the effectiveness of the search system is entirely dependent on the quality of the data, and the logic of the search sequence applied by the user. The biology and distribution fields are qualitative data in the form of extracted texts (see 3.1.2.1), and as such may very well contain text which could create misleading results. Lindroth (1985), for example, writes “Europe: not UK or Iberia” in his distribution notes for *Carabus coriaceus* L. The search system, on being asked to return species where the distribution data mentions ‘UK’ or ‘Iberia’ will return this species, ignoring the ‘not’. An attempt has been made to design an ‘intelligent’ searching system which would be able to handle the ‘not UK’ case, but the ‘not ... Iberia’ is a more complex problem. This system has not been fully implemented in the release associated with this thesis, and users are recommended to always thoroughly examine the raw data for any species lists derived in this way.

Although the search system may appear daunting, it allow complex queries to be performed by linearly adding search terms and logically applying them to the previous search results. One of the more advanced features of the search system is that it is able to retrieve a list of sites from which any of the resulting species are known. Clicking the [Report sites with any of these species] button performs this action, and after a few seconds open a preview of a report which lists the summary data for each site, followed by a list of the species matched between the site list and the search results (see section 3.5.4). In principle, this allows researchers to define a specific environment by the application of search terms, and then retrieve a list of sites where that environment may have been present. This is an invaluable feature for comparative studies, and in general site interpretation. As always, this report may be exported to MS Word or Excel and saved, or printed directly.

Search result species lists can also be exported as Bugs Countsheet formatted MS Excel files, with a single presence-only sample (the final reporting option in the Report Generator, Figure 3.49a). This file can then be re-imported into a BugsCEP site and treated in the same way as any other countsheet.

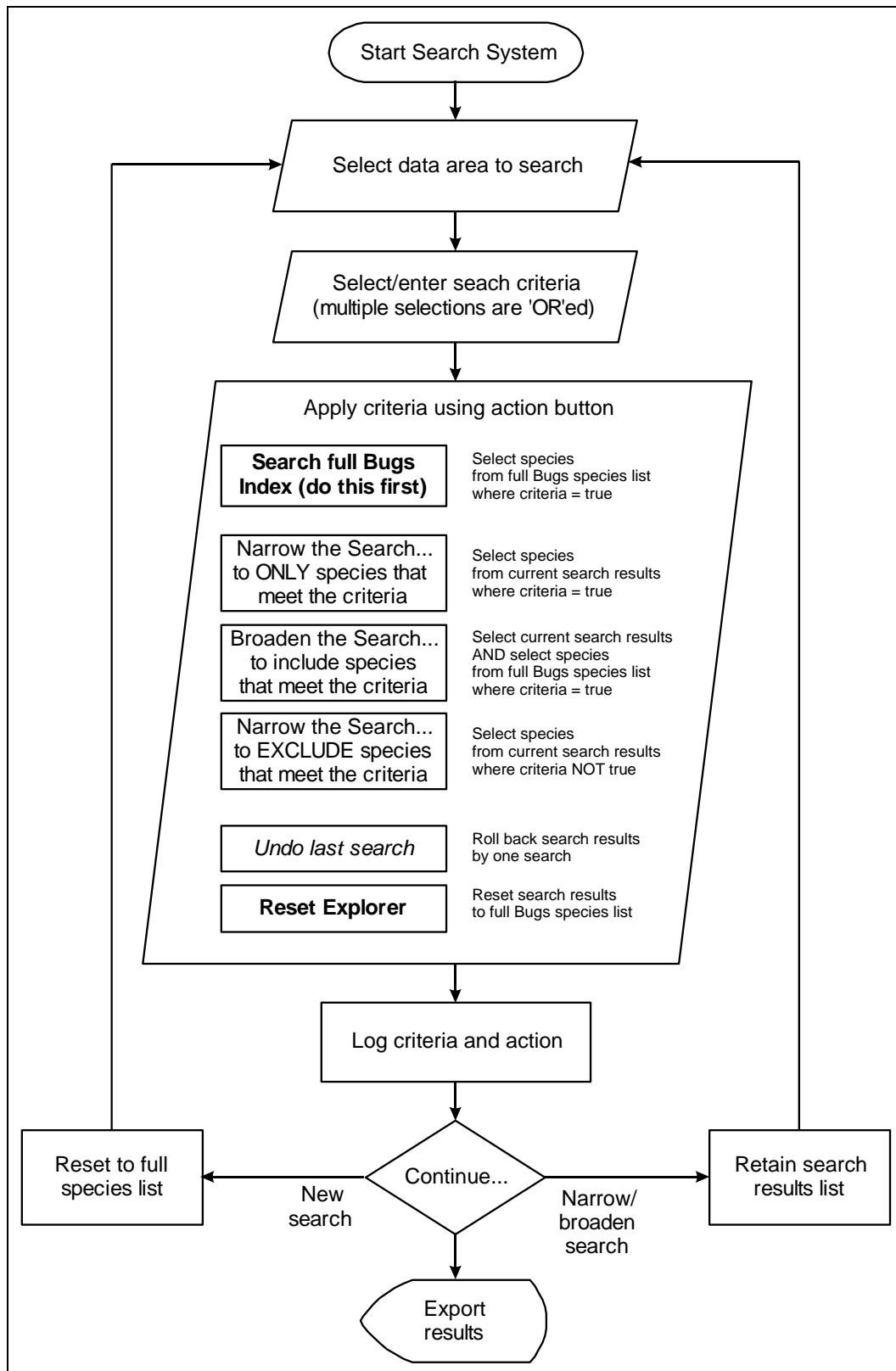


Figure 3.44. Flow diagram illustrating the sequence of events when using the Bugs Search Explorer.

3.4.5.1 A worked example

Imagine the scenario where a researcher wishes to retrieve a list of aquatic beetle species, which are extinct in the UK, but can be found still living in Sweden. The following steps describe how to answer this using the Bugs Search Explorer, with *italics* denoting the start of stages in criteria specification.

1. Reset the explorer, if a previous search is still active, by clicking the [Reset Explorer] button.
2. Select the [RDB] criteria type tab.
3. Select the ‘United Kingdom: UKRDB’ item from the [Select Country & RDB System] drop down box.
4. Click the ‘X: Extinct in UK’ item from the [Select Category] list, and click the [Add Class to Selected Criteria] button to add an ‘X’ to the [Selected Criteria] list.
5. Click the [Search full Bugs Index...] button. The Current Search Results are updated, with only 97 taxa now being displayed – these are all extinct in the UK, but more than just aquatic species are included. An item is also written to the Search Event Log.
6. Select the [Bugs EcoCodes] criteria type tab.
7. Click the ‘Aquatics’ class, and press the [Add Class to Selected Criteria] button to add ‘BEco01’^{xxxiv} to the [Selected Criteria] list.
8. Click the [Narrow the Search... to ONLY species that meet the criteria] button. This results in a list of only ten taxa that are extinct in the UK and classified as aquatic according to the BugsCEP classification system.
9. Select the [Distribution Text] criteria type tab.
10. Type the word ‘Sweden’ into the [Enter single search term:] box, and click [Add entered term to Selected Criteria].
11. Click the [Narrow the Search... to ONLY species that meet the criteria] button. This results in seven taxa, the names of which can be exported by pressing the [Create Search Results/Export] button and choosing the [Just the names] report (Figure 3.45). Page two of the report shows the search event log which lead to the species list (Figure 3.46), allowing the search to be easily replicated. The date of creation of the report is also given, as this could be important when comparing results from different versions of BugsCEP.

The accuracy of this list is dependant on a number of factors (in addition to those mentioned above), including the accuracy and completeness of the data, the reliability of the cited authors, and the reliability and extent of the Bugs EcoCode classifications. As a result the list should in no way be considered a fact, but only a suggestion, and an aid to further research. Future developments of the BugsCEP system will inevitably improve the search system, in particular by the classification of more species, the addition of real geographical distribution ranges and the ability to search fossil records and climate data through this interface.

Bugs Search Explorer Results		
Reference for version of Bugs used for this report (list all components used when referencing):		
Buckland P.I. & Buckland P.C. (2006). Bugs Coleopteran Ecology Package (Versions: BugsCEP v7.61; Bugsdata v7.09; BugsMCR v2.01; BugStats v1.22) [Downloaded/CDROM: DATE] www.bugscep.com.		
Code	Family	Genus & species
4.0270032	DYTISCIDAE	Colymbetes striatus (L.)
4.027004	DYTISCIDAE	Colymbetes dolabratus (Payk.)
4.028002	DYTISCIDAE	Hydaticus continentalis Balfour-Browne
4.029001	DYTISCIDAE	Graphoderus bilineatus (Deg.)
4.032001	DYTISCIDAE	Cybister lateralimarginalis (Deg.)
5.002007	GYRINIDAE	Gyrinus natator (L.)
8.001001	SPERCHEIDAE	Spercheus emarginatus (Schall.)

Figure 3.45. Search results report showing ‘Just the names’ of the taxa which are RDB classified as extinct in the UK, classified as aquatic and known from Sweden.

^{xxxiv} This is the Bugs EcoCode designation for aquatic taxa, as described in section 4.3.1.



Figure 3.46. Final page of the ‘Just the names’ search results report, showing the sequence of events that lead to the final species list (the search log).

3.4.6 The help files

A simple, example based html help system is built into BugsCEP, the contents page of which (Figure 3.47) is accessible through the [BugsCEP Help Contents] item of the Help menu. Context sensitive help, that is to say specific help for the BugsCEP screen that is currently open, is available by pressing the <F1> function key on the keyboard. This opens the appropriate help page for the currently viewed area of the software, but, unfortunately does not yet move to the exact description of the currently selected button or feature. The help pages will be improved with time, and appropriate tool tips, indices and wizards will be added if users request them.

The entire help system may also be navigated through a series of buttons that are shown at the top of every help page (Figure 3.47). The system, which closely resembles a website, is essentially organised by program component, as described above, with additional introductory and overview pages to complement the information. Clicking on the title bar will attempt to open the BugsCEP online website through the help system.

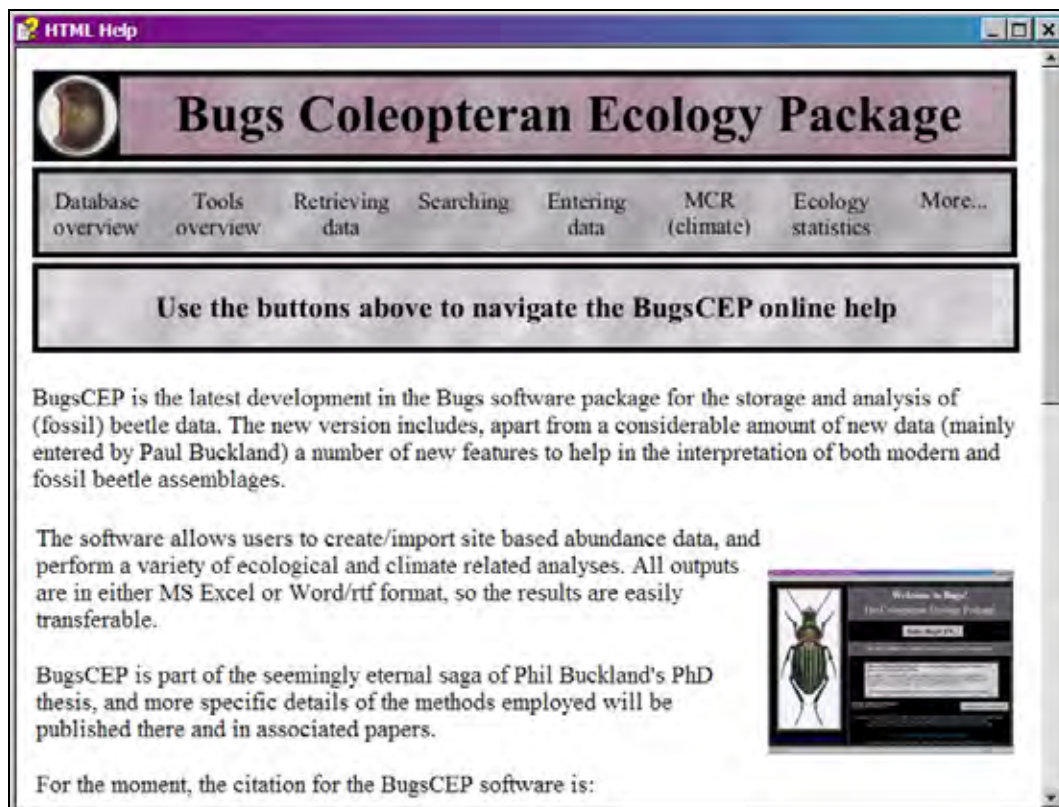


Figure 3.47. Screenshot: The BugsCEP help contents page, the links below the title bar providing navigation.

The help files are routinely duplicated on the website at <http://www.bugscep.com/help/> where they are accessible to anyone with an Internet connection. It often proves useful to be able to direct users to

instructions using a web address rather than trying to describe what to do over the phone or via email. The exposure of the help pages on the Internet also increases the visibility of the website through search engines, which are a valuable form of passive marketing and help to increase awareness of the software.

3.5 Reporting & Exporting Functions

BugsCEP includes a number of export and report functions that are spread throughout the program, which are presented here by program area.

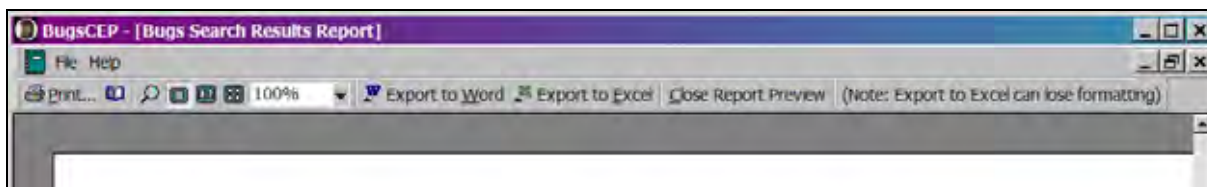


Figure 3.48. Screenshot: The report toolbar, displayed at the top of all report previews.

Whenever a report is created in BugsCEP, it is first displayed in report preview mode, and the standard toolbar and menu is replaced with those shown in Figure 3.48. The first two buttons activate standard Print and Page Setup functions, to the right of which are a number of zoom controls which allow the report to be previewed as up to six pages at once. Following these are the [Export to Word] and [Export to Excel] buttons, both of which export the report to a temporary file (rtf format for MS Word, and xls for MS Excel) and then open it in the indicated program^{xxxv}. The user can then save this file for future use.

A number of functions in BugStats and BugsMCR export complex data or results directly to MS Excel, and these files are described above (sections 3.4.3.3 and 3.4.4) and in Chapters 4 and 5.

3.5.1 Single species data

It is possible to output the information available for a single species by clicking the [Export] button on the BugsCEP main screen whilst viewing the species information. A confirmation request will be displayed, which can be prevented from appearing again, and then the BugsCEP **Report Generator** options pop-up window (Figure 3.49). This pop-up is used in more than one section of the program, and all the available options, which are described in section 3.5.4, may not be useful in every context.

After selecting the required report type, and clicking the [Preview/Export Report] button the appropriate report preview will be shown. The relevant references will be exported with reports (with the exception of XLS file output), and users are reminded to always cite the original data sources *as well* as BugsCEP. A file name will be requested if the [Export as Bugs Countsheet] option is chosen. The fossil record for a taxon cannot currently be exported in this way, and must be manually selected, copied and pasted into other software. A future update will provide this facility.

^{xxxv} Note that the [Export to Excel] function of the Countsheet Manager, or [Export as Bugs Countsheet] report option (Figure 3.49), rather than text reports, are recommended when wishing to export species lists as they provide better formatting.

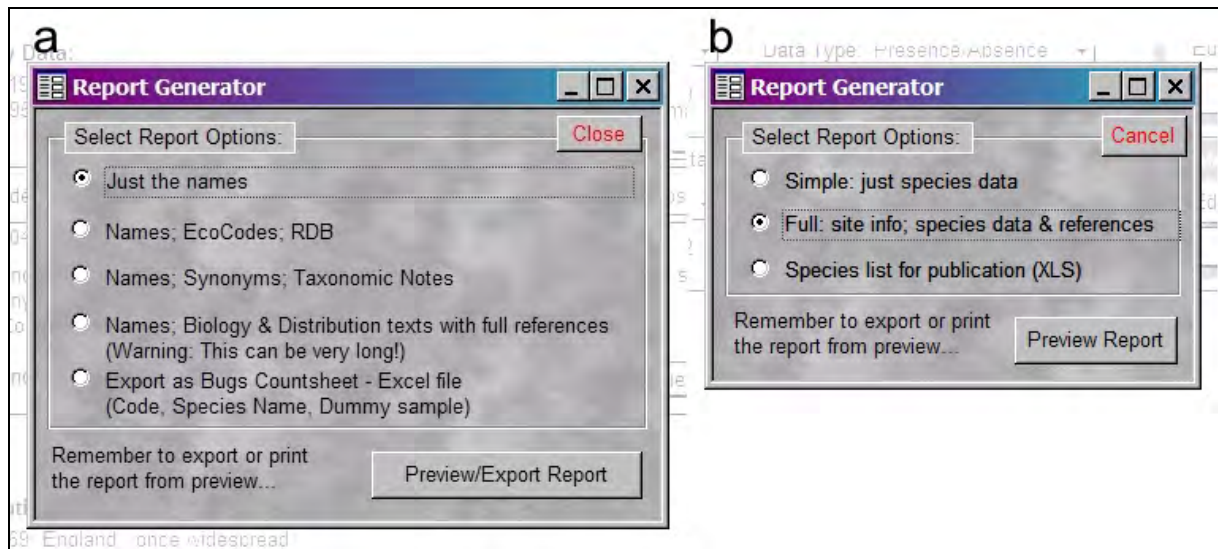


Figure 3.49. Screenshot: The Report Generator, which allows the user to specify which type of report to create. It is used in several areas of the program, and has a number of variations as shown by **a** and **b**.

3.5.2 Bibliographic data

The BugsCEP Bibliographic Browser includes the facility to output the references being displayed at any time to a standard report, by clicking the [Export current references] button. This report simply lists the active references by author.

3.5.3 Site reports

A compilation of the biology and distribution data for the species found at a site, along with the relevant references, can be created by clicking the [Create Report] button on the Countsheets Manager screen for any site. The Report Generator will open and display three report options (Figure 3.49b), the 'Full...' report, being the default. This report includes site information, including site references, followed by the biology and distribution information for each taxon found at the site, along with that taxon's total abundance (Figure 3.50). References to all works cited are exported at the end of the report.

BugsCEP Full Site Report

Reference for version of Bugs used for this report (list all components used when referencing):
 Buckland P.J. & Buckland P.C. (2008). Bugs Coleopteran Ecology Package (Versions: BugsCEP v7.61; Bugsdata v7.09; BugsMCR v2.01; BugStats v1.22) [Downloaded: C:\CDROM: DATE] www.bugscep.com.

<p>Site Name: Saint Bees Region: Cumbria Country: England</p>	<p>National Grid: NX 965114 Latitude: 54°29'14"N Longitude: 3°35'57"W Altitude:</p>
--	--

Identifications by: Pearson, Coope & Joachim
 Specimens:

Lateglacial kettle hole in sea cliffs section, originally by Pearson.
 See restudy by Coope & Joachim (1980).

Site References

Coope, G.R. & Joachim, M.J. (1980)
 Lateglacial environmental changes interpreted from fossil Coleoptera from St Bees, Cumbria, N.W. England. In J.J. Lowe, J.M. Gray & J.E. Robinson (eds.) Studies in the Lateglacial of North-West Europe, 55-68. Pergamon, Oxford.

Pearson, R. G. (1962)
 The Coleoptera from a Late Glacial Deposit at St. Bees, West Cumberland. Journal of Animal Ecology, 31, 129-150.

Countsheet: Saint Bees Coope bugsdata.XLS

Species Biology and Distribution data

1.0010122 <i>Cicindela</i> sp.	Abundance at site: 1
1.0040100 <i>Carabus prohemeranus</i> Hbst.	Abundance at site: 3

Beaghton 1951: on dwarf shrub (*Calluna*) heaths, also in infield and outfield grassland.
 Coulson et al. 1955: in pitfall traps in Flow country;
 den Boer 1977: dry forests, also in coniferous plantations; reproduces in autumn; larvae surface active in winter & early spring. Brachypterous.
 Duff 1993a: under logs or stones in woods and grassland.
 Duff & Luff 1990: in 3 chalk grassland, with *L. depressus* & *A. aenea*.
 Eyras 2000: prey includes Lumbricidae, enchytraeidae, Nematoda, Diplopoda, Acari, Collembola, Diptera, and Coleoptera; will also eat fungi. May consume dead invertebrates.
 Forsythe 1952: predominantly a predatory surface runner; gut analyses show large intake of slug tissue.
 Harde 1954: June to September, the commonest *Carabus* on moor and heaths in Britain.
 Larsson & Gajda 1979: subsp. *Carabus p. fennicus*. Under stones, on meadows, moist biotopes, often near watercourses. In Scandinavia xerophilous, on gravelly soil.
 Calluna heaths. In C. Europe in light woodland.
 Lindroth 1974: subsp. *Carabus p. gallicus varietas v. procedens*. Open dry country; light woodland, mostly heath, Highland.
 Lindroth 1985: nocturnal, several subsp. *Carabus p. gallicus*, *C. p. moeki*, *C. p. strandi*. Prefers warm loc. open, dry biotopes; gravelly; sandy soil. moraine. On *Calluna* heaths, pine woodland in Denmark in light deciduous woodland. Biennial. Adults late summer to autumn.
 Luff 1981: Key to separate form *C. violaceus*.
 Richards 1926: flightless, typical of *C. fennicus*.

Woodcock 2001
 Woodcock, B. (2001). Some uncommon Staphylinidae (Col.) from South Aberdeen shire (VC92) including *Eurypterus picipipes* (Paykull). Entomologist's monthly Magazine, 137, 244.

Young 1977
 Young, M. R. (1977). A preliminary report on the distribution of *Gyrinus opacus*. Balfour-Browne Club News letter, 4, 12, 14-15.

Zimmerman 1980
 Zimmerman, J.R. (1980). Use of multivariate procedures in studies of species problems in the sculptilis group of North American Colymbetes (Coleoptera: Dytiscidae). Coleopterists Bulletin, 34, 213-226.

Zimmerman 1981
 Zimmerman, J.R. (1981). A Revision of the Colymbetes of North America (Dytiscidae). Coleopterists Bulletin, 35, 1-62.

Figure 3.50. Site report, full version, showing part of the first and last pages of the 82 page report generated for the Saint Bees site. The first page shows site summary information and the first taxa from the site, and the last page shows the last few references cited in the biology and distribution data.

3.5.4 Reporting search results

A number of reporting options (Figure 3.49a) are available when a list of species is arrived at through the Bugs Search Explorer, each of which can be created by pressing the [Create Search Results Reports/Export] button. The figures shown below are all produced from the results of the worked example given in section 3.4.5.1. The search event log (see Figure 3.46) is appended to all reports just before the list of references, and will not be duplicated below.

The simplest report form, 'Just the names', requires little explanation and is shown above (Figure 3.45 and Figure 3.46). The 'Names; EcoCodes; RDB' report lists the Bugs ecology classifications, and Red Data Book classifications for each species (Figure 3.51).

Bugs Search Explorer Results		
Reference for version of Bugs used for this report (list all components used when referencing):		
Buckland P.I. & Buckland P.C. (2006). Bugs Coleopteran Ecology Package (Versions: BugsCEP v7.61; Bugsdata v7.09; BugsMCR v2.01; BugStats v1.22) [Downloaded/CDROM: DATE] www.bugscep.com.		
Note: Referencing the correct Bugs Stats (EcoCodes) version allows intercomparability of results.		
Code	Family	Genus & species
4.0270032	DYTISCIDAE	Colymbetes striatus (L.)
	BugsEcoCODE: BEco01a: Standing water (ponds) indicators	
	BugsEcoCODE: BEco01: Aquatics	
	Rarity: United Kingdom	X: Extinct in UK
		UKRDB
4.027004	DYTISCIDAE	Colymbetes dolabratus (Payk.)
	BugsEcoCODE: BEco01a: Standing water (ponds) indicators	
	BugsEcoCODE: BEco01: Aquatics	
	Rarity: United Kingdom	X: Extinct in UK
		UKRDB
4.028002	DYTISCIDAE	Hydaticus continentalis Balfour-Browne
	BugsEcoCODE: BEco01a: Standing water (ponds) indicators	
	BugsEcoCODE: BEco01: Aquatics	
	Rarity: United Kingdom	X: Extinct in UK
		UKRDB
4.029001	DYTISCIDAE	Graphoderus bilineatus (Deg.)
	BugsEcoCODE: BEco01a: Standing water (ponds) indicators	
	BugsEcoCODE: BEco01: Aquatics	
	Rarity: United Kingdom	X: Extinct in UK
		UKRDB
4.032001	DYTISCIDAE	Cybister lateralimarginalis (Deg.)
	BugsEcoCODE: BEco01a: Standing water (ponds) indicators	
	BugsEcoCODE: BEco01: Aquatics	
	Rarity: United Kingdom	X: Extinct in UK
		UKRDB
5.002007	GYRINIDAE	Gyrinus natator (L.)
	BugsEcoCODE: BEco01a: Standing water (ponds) indicators	
	BugsEcoCODE: BEco01: Aquatics	
	Rarity: United Kingdom	X: Extinct in UK
		UKRDB

Figure 3.51. Search results report using the ‘Names; EcoCodes; RDB’ option. Each taxon name is followed by its BugsCEP habitat and Red Data Book classifications.

The ‘Names; Synonyms; Taxonomic Notes’ option may be more useful for taxonomic studies, or when identifying specimens, although these field contain a limited amount of information at the present time (Figure 3.52). For many users, the ‘Names; Biology & Distribution texts with full references’ option will be the most useful, as it exports the full extent of the databases extracted texts for each species, along with references to the cited works (Figure 3.53).

Bugs Search Explorer Results		
Reference for version of Bugs used for this report (list all components used when referencing):		
Buckland P.I. & Buckland P.C. (2006). Bugs Coleopteran Ecology Package (Versions: BugsCEP v7.61; Bugsdata v7.09; BugsMCR v2.01; BugStats v1.22) [Downloaded/CDROM: DATE] www.bugscep.com.		
Code	Family	Genus & species
4.0270032	DYTISCIDAE	Colymbetes striatus (L.)
4.027004	DYTISCIDAE	Colymbetes dolabratus (Payk.)
4.028002	DYTISCIDAE	Hydaticus continentalis Balfour-Browne
	Synonym: Hydaticus stagnalis (F.) non (Fourcroy)	
	Duff 2006	
	Taxonomy notes: Hodge & Jones 1995: very similar to H transversalis, but black area on hind margin of pronotum less extensive and elytra with longitudinal streaks; both characters variable.	
4.029001	DYTISCIDAE	Graphoderus bilineatus (Deg.)
4.032001	DYTISCIDAE	Cybister lateralimarginalis (Deg.)
5.002007	GYRINIDAE	Gyrinus natator (L.)
	Taxonomy notes: Foster 2006: Separated from Gyrinus substriatus on female genitalia.	
	Taxonomy notes: Angus & Carr 1982: Detailed separation from Gyrinus substriatus.	
8.001001	SPERCHEIDAE	Spercheus emarginatus (Schall.)

Figure 3.52. Search results report using the ‘Names; Synonyms; Taxonomic Notes’ option.

A list of the sites from which the species are known, along with site summaries is available by clicking the [Report sites with any of these species] button on the Search Explorer. The first few items in the report for the seven species listed above are shown in Figure 3.54.

Bugs Search Explorer Results		
Reference for version of Bugs used for this report (list all components used when referencing):		
Buckland P.I. & Buckland P.C. (2006). Bugs Coleopteran Ecology Package (Versions: BugsCEP v7.61; Bugsdata v7.09; BugsMCR v2.01; BugStats v1.22) [Downloaded/CDROM: DATE] www.bugscep.com.		
Code	Family	Taxon
4.0270032	DYTISCIDAE	Colymbetes striatus (L.)
<p>Detmer 1983: Key to C spp. Koch 1989: in bog pools. Nilsson & Holmen 1995: in more or less temporary ponds in open country; in S, more stable populations usually in smaller dystrophic ponds or clear water lakes or in ponds and ditches in bogs and fens. Frequently on flooded lake or river margins. Adults overwinter and larvae in late spring and early summer.</p> <p>Detmer 1983: Europe: N & parts of C, Russia, Siberia. Koch 1989: Europe: E. Nilsson & Holmen 1995: Denmark: scattered rec in E. Norway: mainly in SE, but also in ST. Sweden & Finland: widespread & common, N to Arctic Circle. Baltic States. W Palearctic: E France & Switzerland to Fennoscandia, E to W Siberia.</p>		
4.027004	DYTISCIDAE	Colymbetes dolabratus (Payk.)
<p>Böcher 1988: Greenland: typically in ponds with rich vegetation, highly eurytopic in almost any kind of stagnant water, except possibly temporary pools. In summer, adults and larvae found together, but in late summer and autumn the imagines migrate to hibernate in lakes which do not freeze to the bottom, i.e. > 3m. Flight records from Scoresby Sound. In Scandinavia, alpine and subalpine, found in marshes, sphagnum bogs and ponds devoid of vegetation. On Hardangervidda, Norway, almost confined to pools above 1000m, without vegetation, but with Carex spp, grasses and mosses along shore. Feeds probably on crustacea, Brachinecta, Cladocera. Adults throughout year. Pupa is terrestrial, found in small hollows under moss close to shore of ponds.</p> <p>Larsson & Gjøga 1959: - Iceland - throughout but rarer in S. and S.W.; prefers ponds rich in plants, but often in other biotopes, even clayey pools near glacier's edge.</p> <p>Lindroth et al. 1973: in ponds, usually with rich vegetation. Wings full and functional. Icelandic form has been given the subsp name Thomsoni Sharp.</p> <p>Nilsson & Holmen 1995: mainly in shallow ponds above the treeline, frequent on flooded lake and river margins, also in ponds in gravel workings. Larval development mainly late June-late July. Adults ready fliers, in Greenland migrate to deeper lakes for overwintering.</p> <p>Zimmerman 1981: a highly variable sp. collected in pools of water on pack ice!</p> <p>Böcher 1988: Greenland: most common and best known beetle, both coastal and inland from Kap Farvel in S to Upemavik on W and Clavering O on E. N limit is set by 10 month of freezing on lakes because of need to hibernate. Circumpolar, boreal to arctic, Europe, Iceland, N Scandinavia, N Russia, N Siberia, E to Kamchatka, Alaska, Canada: NW Territories to Labrador.</p> <p>Larsson & Gjøga 1959: Iceland.</p> <p>Lindroth et al. 1973: Circumpolar N, Greenland, Iceland: throughout but rare in S.</p> <p>Nilsson & Holmen 1995: Fennoscandia: mountains S to TE in Norway & DR to Sweden. Finland: N, S to Ob. A circumpolar sp, in N America along N edge of continent from Alaska to Quebec. Greenland N to 75N. Eurasia, through Arctic and N boreal regions from Iceland and</p>		

Figure 3.53. Search results report using the ‘Names; Biology & Distribution texts with full references’ option.

Bugs Search Explorer Results - Sites where selected taxa found	
Reference for version of Bugs used for this report (list all components used when referencing): Buckland P.I. & Buckland P.C. (2006). Bugs Coleopteran Ecology Package (Versions: BugsCEP v7.61; Bugsdata v7.09; BugsMCR v2.01; BugStats v1.22) [Downloaded/CDROM: DATE] www.bugscep.com.	
Ageröd V, Skåne, Sweden	
Ageröd V. Specimens hand picked by the excavators of a mesolithic midden. Lat: 55°55'60"N Long: 13°23'60"E Altitude: 39m Identified by: Lemdahl	
<i>Search Result species found at site:</i> 4.0320010 DYTISCIDAE <i>Cybister lateralimarginalis</i> (Deg.)	
Björkeröds Mosse, Skåne, Sweden	
Lateglacial to Early Holocene successions from profiles A6, B1 & 3b. Dates from base of A6, the remainder interpolated from pollen data. Note : Sample A6/1 is Early Holocene (EH). Lat: 56°16'60"N Long: 12°32'0"E Altitude: 75m Identified by: Lemdahl	
<i>Search Result species found at site:</i> 4.0270040 DYTISCIDAE <i>Colymbetes dolabratus</i> (Payk.)	
Brandon Upper Deposit, Warwickshire, England	
Initially interpreted as Mid-Weichselian organic sediments, but see Maddy et al. (1994) for suggestion that this and Upper 2 are OS 6. Lat: 52°27'0"N Long: 1°23'0"W NGR: SP391753 Identified by: Coope Specimens stored at University of Birmingham, UK.	

Figure 3.54. Report showing the first few sites from which the seven species shown in Figure 3.45, which were retrieved through the Search Explorer, are known.

3.6 The Addition of Taxa, Biology/Distribution Data and References

Modern reference data entry is currently undertaken centrally by Paul Buckland, and the lack of update routines for the import of data entered by other users makes decentralization potentially difficult and time consuming. The current system also allows for more uniform implementation of data quality and validation criteria. Users are instructed to contact the authors with requests for the addition of specific data. Should additional taxa or reference data be required, then the authors would be only too pleased to add them on request. Users may, on the other hand, add as many countsheets and sites as desired. References for sites may be added by clicking the [Add Reference] button, on the Site Information screen. This becomes available either on the creation of a site, or when the [Edit Site] button is pressed and the correct 'Admin' level password entered. The reference entry system automatically inserts lettered suffices where duplicate author-date combinations occur (e.g. 'Buckland 2000a', 'Buckland 2000b', etc.), and allows the user to pick any existing reference for use in the current context.

A number of maintenance features have been omitted from this text as they would have been out of place, and will be documented at a later date. These allow data managers complete control over the taxonomic master list, ecology code and other lookup systems, as well as providing fail safes for when data become corrupt.

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-expertⁱ 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 scleritesⁱⁱⁱ, 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 it 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 known 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 led 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.
<i>BEco1a</i>	<i>Indicators: Standing water</i>	<i>Primary habitat in/on pools, ponds, slow flowing water – including temporary ponds, but not species specifically in vegetation and mud or banks of ponds.</i>
<i>BEco1b</i>	<i>Indicators: Running water</i>	<i>Rivers and/or streams. Species predominantly found in these.</i>
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.
<i>BEco4a</i>	<i>Indicators: Deciduous</i>	<i>Specifically deciduous wood or woodland, species not found on coniferous wood except on rare occasions.</i>
<i>BEco4b</i>	<i>Indicators: Coniferous</i>	<i>Specifically coniferous wood or woodland, species not found on deciduous wood except on rare occasions.</i>
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.
<i>BEco7c</i>	<i>Indicators: Dung</i>	<i>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.</i>
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 section 4.3.2). 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 illustrate 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
<i>Otiorhynchus</i> sp. (29 spp.)	BEco4; BEco3; BEco5a; BEco6a; BEco6b; BEco12; BEco9a; BEco13
<i>Agabus</i> 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.

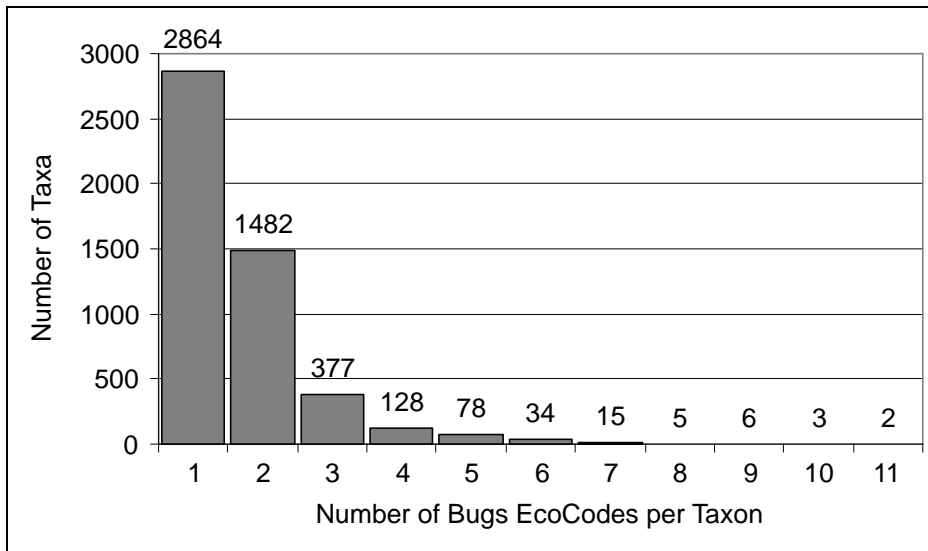


Figure 4.1. Number of Bugs EcoCodes per taxon.

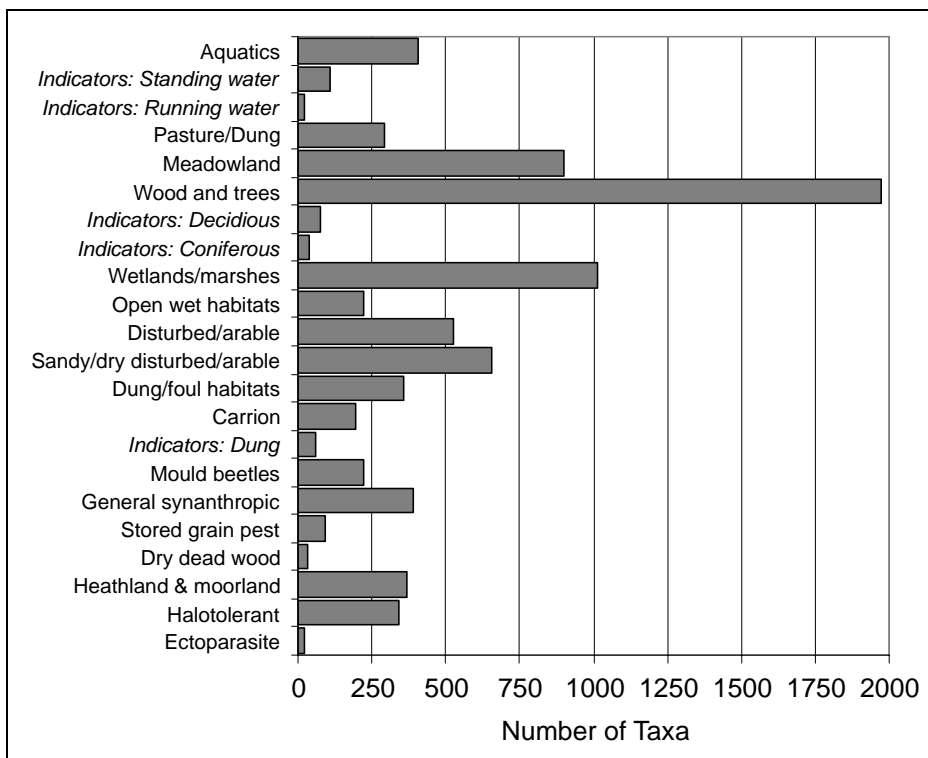


Figure 4.2. Number of taxa per Bugs EcoCode habitat class. Taxa can be present in more than one general class, but only one ‘Indicator’ class.

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.

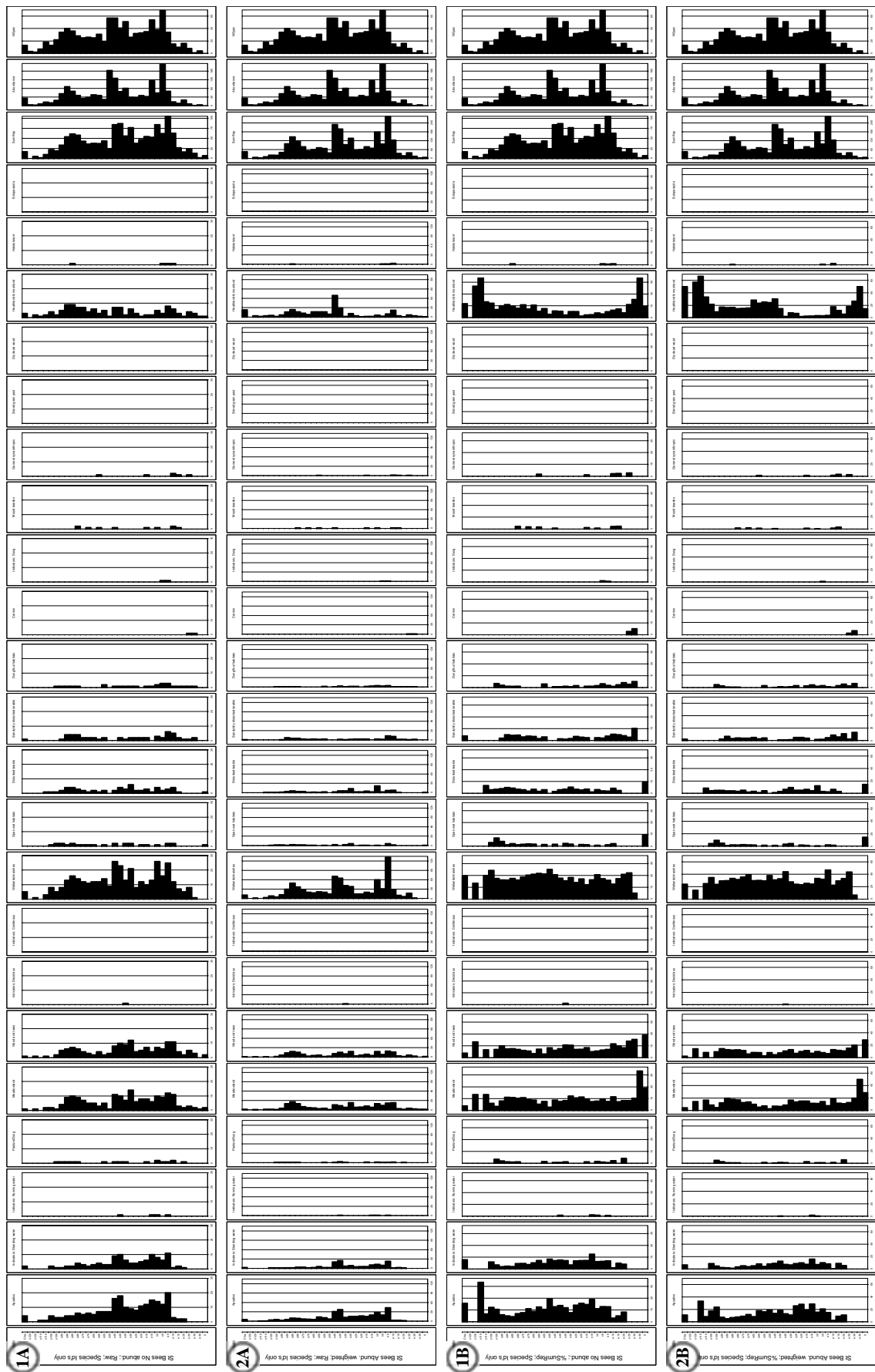


Figure 4.3. *Unmodified* BugStats EcoFigs for Saint Bees (Coope & Joachim, 1980), note that the basal sample is at the top of each diagram. Settings, from the left: 1A: No abundance, Raw; 2A: Abundance weighted, Raw; 1B: No abundance, %SumRep; 2B: Abundance weighted, %SumRep (see 4.3.3.3).

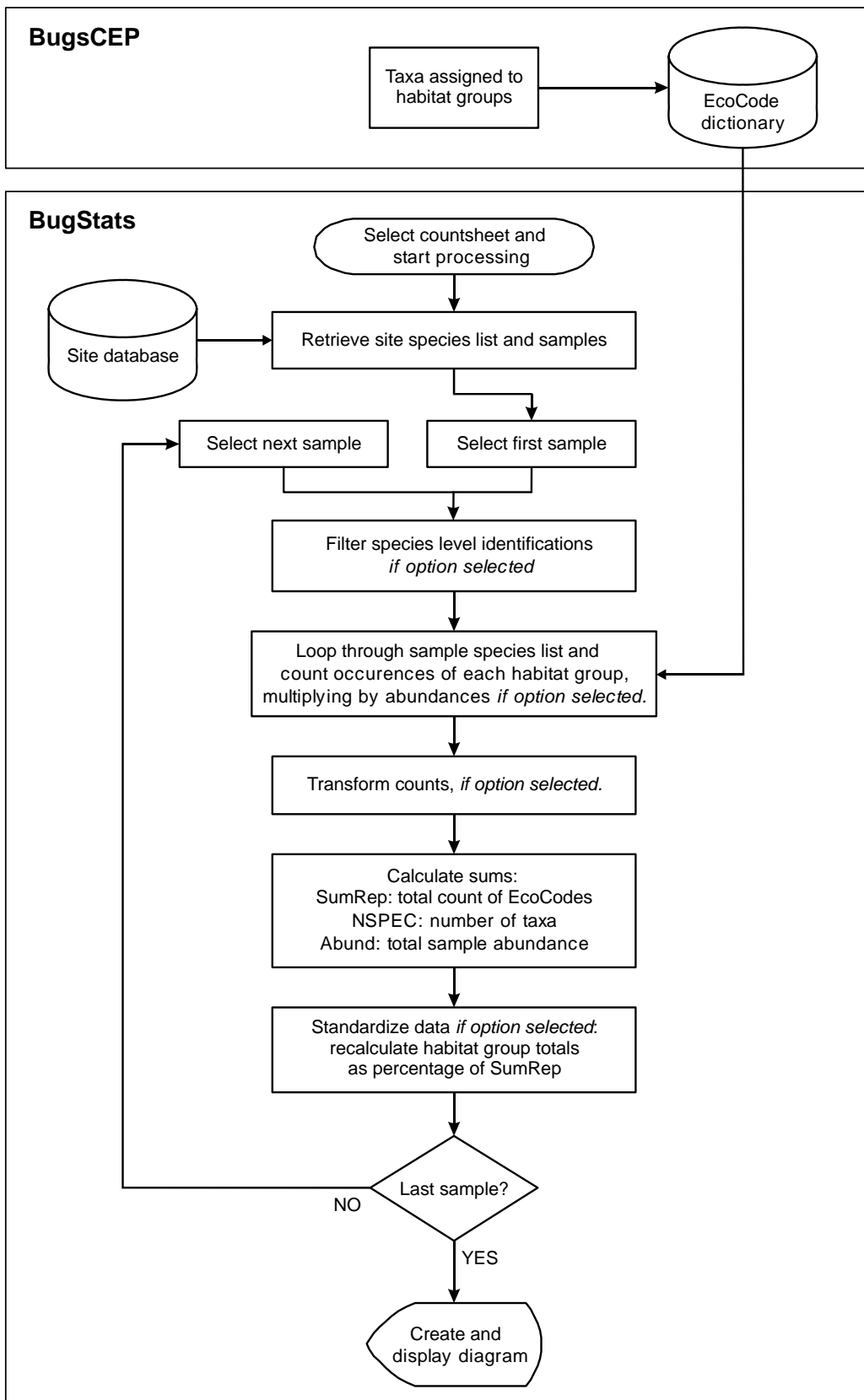


Figure 4.4. Flow diagram illustrating the sequence of events that create an EcoFig diagram from site data.

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.

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

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.

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

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 be 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, for 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	phyllo-detriticolous	
	boleticolous	lignicolous	phyto-detriticolous	torrenticolous
	cadavericolous	limnicolous	planticolous	xylo-detriticolous
	campicolous	linicolous	polyporicolous	zoo-detriticolous
	cavernicolous	microcavernicolous	praticolous	
	corticolous	muscolous	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		
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.

5 BugsMCR: Software for MCR Temperature Reconstruction from Beetle Assemblages

This chapter describes the BugsMCR climate reconstruction component of the BugsCEP software. Some developmental aspects are described, such as the improvements in BugsMCR over previous software, but the primary focus is on the implementation of the methodology itself, its advantages and disadvantages, and possible improvements. Instructions on the use of BugsMCR are given in section 3.4.4, and some practical applications can be found in Chapter 6. Whilst the MCR, and prediction by MCR, program modules are fully functional in the current release of BugsCEP, some of the more advanced techniques described below are not publicly available and will be released in conjunction with later publications. They are presented here to prepare the ground and to enable the discussion of future developmental strategies.

Fossil beetle assemblages have been used for the reconstruction of Quaternary palaeoclimates since the late 1950s, when Russell Coope first used them to reconstruct late Quaternary environments at Chelford in Cheshire (Coope, 1959) and Upton Warren in Worcestershire (Coope *et al.*, 1961). The Mutual Climatic Range method (MCR) (Atkinson *et al.*, 1986) was developed to provide quantitative climate reconstructions, using the overlap of modern species thermal envelopes to predict/retrodict the temperature regime in which all species in a fossil assemblage would have been able to survive. During the late 1980s and early 1990s, two software suites were developed for the generation of thermal envelopes and calculation of MCR, commonly referred to as MCRBirm and MCRUEA after their respective development institutions (Birmingham University and the University of East Anglia). The reconstruction components of these software, called RECON and RECON2 respectively, run in the MS-DOS environment, and there has for some time been the need for a more user-friendly format using MS Windows, which is now provided in BugsMCR.

Much has been written on the MCR method and its applications (see references throughout this chapter), although detailed descriptions of its development and mechanics are rare, and almost entirely restricted to unpublished PhD theses (see section 5.1.1).

5.1 Background and Software Development

Although detailed descriptions of the earlier MCR software are hard to find (e.g. Perry, 1986), the MCR method itself is described in a number of articles (e.g. Atkinson *et al.*, 1986; 1987; Lowe & Walker 1997). BugsMCR was initially developed, by the author of this thesis, to improve upon the original MS-DOS software, as a standalone program for use in research and teaching, and has been subsequently incorporated into the latest version of BugsCEP, with a number of improvements. This allows for a more integrated approach to fossil beetle studies, combining climate change with ecology and biogeography. Sites can easily be analysed using the full suit of BugsCEP tools without the need for transferring data between programs.

5.1.1 The mutual climatic range (MCR) method in brief

The mutual climatic range is the term used to define the set of temperature valuesⁱ in which a group of beetles, or indeed any organism, can all survive. The method overlays the known thermal tolerances, or thermal envelopes, of the taxa in a sample and simply reads off the extremes of the area of greatest overlap (Figure 5.1). The greatest overlap is typically expressed as a percentage of the number of MCR species, that is to say the species with thermal envelopes, in the sample. By repeating this process for samples in a stratigraphic sequence it is possible to gain a picture of climate change over

ⁱ Additional climate variables have been used in the application of the MCR method to other proxies, e.g. Sinka & Atkinson (1999).

time, should there be sufficient dating evidence to support it. Examples are provided in Chapter 6 of this thesis.

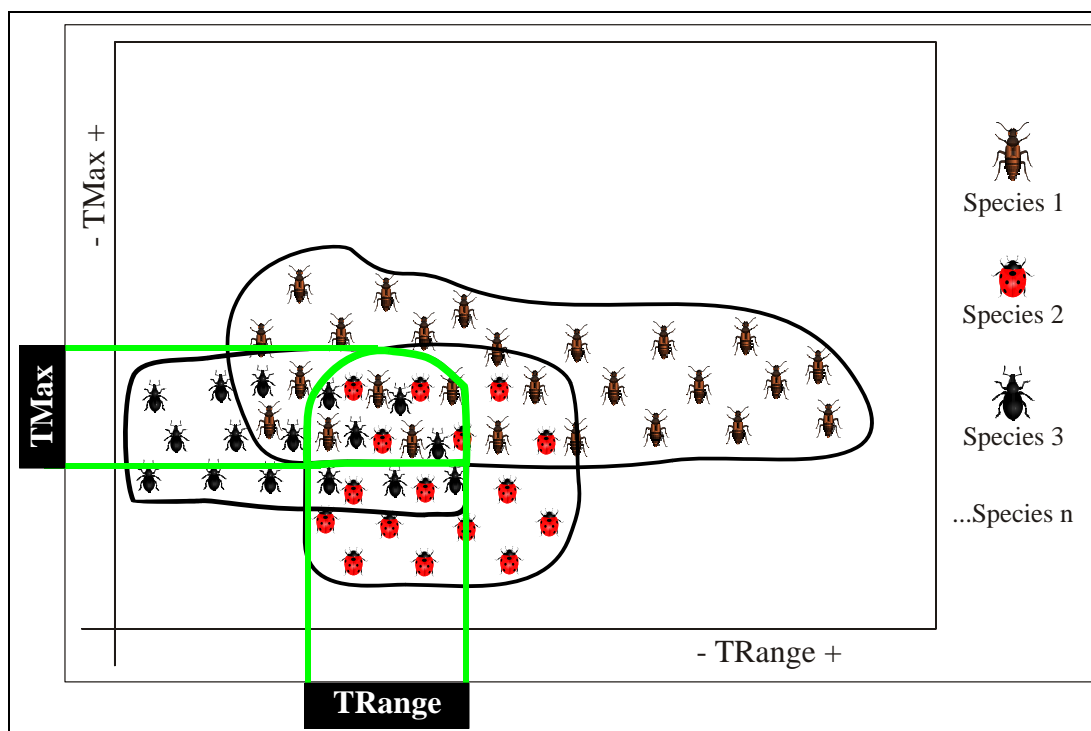


Figure 5.1. An illustration of the derivation of temperature values (TValues) using the MCR method, showing the mutual climatic range (MCR) for three species in climate space.

The method may be classed as an ‘indicator-species approach using two or more climatic variables’ according to Birks (1981), and could be said to have its roots in the works of Iversen (1944), and derivatives, looking at the climatic indicator value of a limited number of plant speciesⁱⁱ. Much of the theoretical, mathematical and computational groundwork for MCR was undertaken in the 1970s and 1980s as components in palaeoentomological PhD theses at Birmingham, UK (Morgan, 1970; Joachim, 1978; Moseley, 1982; Perry, 1986). Ordination was used to show that many beetle distributions are primarily controlled by summer temperatures, and the difference between the summer and winter temperature – summarized in the MCR dataset as TMax: mean temperature of the warmest month, and TRange: difference between TMax and the mean temperature of the coldest month (TMin)ⁱⁱⁱ. TRange is essentially an index of continentality, in that more continental climates show a greater difference between the extremes of summer and winter. The envelopes are stored as TMax versus TRange, and TMin can be calculated by deducting TRange from TMax for any point in TMax/TRange climate space. The reference dataset uses a 1°C cell matrix, making these calculations relatively simple, and leading to reconstructions with a resolution of 1°C in all three thermal dimensions.

MCR relies on the tendency of mobile populations to migrate in the face of climate or environmental change, rather than genetically adapt to the changes and remain in the same location. Remaining in the same location is not an option when the environment changes dramatically over a short period of time, and insects, especially those with wings, are able to react rapidly to abrupt changes (Coope, 1978). Thus extinctions throughout the Quaternary have been few, and the anthropogenic habitat

ⁱⁱ It must be noted that Birks, for a number of reasons that are discussed below, appears to ascribe little value to the MCR method as a quantitative reconstruction technique.

ⁱⁱⁱ Note that occasionally authors (e.g. Huppert & Solow, 2004) have mistakenly cited the use of July and January temperatures instead of the warmest and coldest months. The difference is especially important when considering past shifts in continentality or seasonality, where the warmest and coldest months may change.

fragmentation and destruction of recent centuries may prove to have been more damaging than any other event in prehistory (Thomas *et al.*, 2004)^{iv}. Insects are also among the first organisms to colonise newly exposed land, such as in front of a retreating glacier, or the shores of a drying lake or sinking sea level (see Ashworth, 2001, for a detailed analysis of responses to climate change). A number of omnivorous, aquatic and predatory beetles are particularly mobile and make up a large part of the MCR dataset. It is this rapid colonization ability that first gave proxy evidence for the rapid amelioration at the end of the last Ice Age (Devensian) (Osborne, 1972), and has subsequently revealed a number of other rapid climate change events that floral data, hampered by slower colonization rates, could not resolve. This has led to a significant steepening of the established climate curves, and an intense debate between beetle and chironomid people on the one hand and pollen people on the other, which was only really resolved by the emergence of the Greenland Ice Core data (Alley *et al.*, 1993) showing a similar, if not *more* dramatic set of changes.

5.1.2 The BugsMCR implementation

The usage of BugsMCR has already been described in section 3.4.4, and readers should refer to that section for instructions. The definitions of the variable which specify the reconstructed temperature ranges are repeated here in Table 5.1 for convenience. Although BugsMCR uses the same reference dataset as the Birmingham RECON software, that is where the similarities end. RECON is a piece of compiled FORTRAN software functioning in an MS-DOS environment, and rather than attempting to reverse engineer the source code, it was decided to implement the MCR method from the bottom up using MS Access and VBA. In this respect, although BugsMCR is a natural progression from RECON, it owes nothing to it programmatically. BugsMCR was developed as part of the groundwork for this thesis, and the time for development has been somewhat limited due to other priorities. As a result, some of the useful functions provided in RECON (and its associated programs) have not been replicated in BugsMCR. In particular, it does not provide the facility to view the individual species envelopes, or manually build up overlaps by selecting species. This was a useful investigative and teaching facility included in the MCRUEA software ‘NOD’. Another RECON feature not yet implemented in BugsMCR is that of calculating the warm/cold components of assemblages which do not display a significant overlap. This feature, although it has some theoretical problems, is a useful investigative tool and is considered in more detail below (5.2.5).

Table 5.1. Explanation of MCR results terms, repeated from Chapter 3. See chapters 3 and 6 for worked examples and outputs. The reconstructed temperature range limits are collectively referred to as TValues.

Column	Explanation
Sample	Sample name from database
TMaxLo	Lower limit of reconstructed mean temperature of warmest month
TMaxHi	Upper limit of reconstructed mean temperature of warmest month
TMinLo	Lower limit of the reconstructed mean temperature of the coldest month
TMinHi	Upper limit of the reconstructed mean temperature of the coldest month
TRangeLo	Lower limit of the reconstructed mean temperature ranges (TMax – TMin)
TRangeHi	Upper limit of the reconstructed mean temperature ranges (TMax – TMin)
NSPEC	Number of taxa used in reconstruction
Overlap	Percentage of sample taxa in the area of maximum overlap, used to calculate the temperature values

BugsMCR utilizes the site/countsheets database in BugsCEP, and any site in the database can be processed towards climate reconstruction. This is not to say that all sites are suitable for MCR, and archaeological deposits, in particular with elements thermally cushioned by man-made heat islands,

^{iv} Wikipedia also includes an excellent overview of the ongoing Holocene extinction event: http://en.wikipedia.org/wiki/Holocene_extinction_event

should be treated with extreme caution (see 5.2.2). Abundance data should be entered or imported into BugsCEP as described in section 3.4.2, before they can be accessed through the BugsMCR program component.

The use of MS Access theoretically allows several otherwise complicated calculations and data manipulations to be performed easily and quickly through the use of database queries (cf. stored procedures in SQL-Server), but the data intensive nature of MCR calculations proved incompatible with this technique due to stability and file bloating problems^v. A set of routines was subsequently developed to perform all calculations using arrays in VBA, from which results could be exported to MS Excel for graphing and subsequent user access. In comparison to the previous MS-DOS programs, the facilities for developing a Graphical User Interface (GUI) within MS Access allowed for a more user friendly and intuitive system to be developed (see section 3.4.4 for screenshots). In addition, the use of MS Excel as the output format allowed for easy graphing and compatibility with other statistical packages, although some of the text file outputs of the older versions were more portable. MS Excel, however, is infinitely more flexible in terms of what can be put into the output files, and the current version includes graphing of reconstructed temperatures by sample, the export of raw results, thermal envelopes and species lists, as well as more advanced statistics in versions currently under development (see section 5.1.2.4). MS Excel is also easily integrated with MS Access and VBA, being fully object orientated, and is ubiquitous in academic circles. This enables users to perform subsequent statistical and graphical analyses easily, including the calibration/correction of temperature values if so desired (although see section 5.2.3 on the problems with calibration).

5.1.2.1 *Species thermal envelopes*

The thermal envelope data have been imported and converted from the existing MCRBirm software, and consist of binary grids of 1°C presence/absence data, which were stored in a file called 'beetle.dat'. Figure 5.2 shows the thermal envelope for *Carabus problematicus* Hbst. as stored in MCRBirm (beetle.dat) and BugsCEP respectively. The import process was not without its problems, and due to taxonomic revisions or variations in spellings, 93 of the original taxa had to be manually matched with BugsCEP taxa. The full list of MCR taxa, with their BugsCEP and MCRBirm names can be seen by clicking the [Show All MCR Species](#) link in the BugsMCR interface. This list also includes the numbers used in input files in the original RECON component of the MCRBirm software (Figure 5.3). The latter are provided for backward compatibility and, along with details of the name conversions, to allow for cross-checking of any unexpected problems in the BugsMCR reconstructions.

Although not of the resolution commonly claimed in other climate reconstruction methods, the 1°C grid has proven a reliable system for estimating past temperatures and has been tested against modern, external data sets (Coope *et al.*, 1998). There are currently 436 taxa in the BugsMCR calibration set, although there are plans to increase and refine this, eventually including other climatically sensitive insect groups such as the caddis flies (Trichoptera) (Greenwood *et al.*, 2003). BugsMCR is programmed modularly, the calculation routines and functions are held in modules independent of the interface forms, and can be called from any part of the program. They can thus be easily upgraded should new routines be needed.

^v This is a problem peculiar to MS Access, and the original routines have been preserved for later use in an SQL-Server type version.

CODE	FAMILY	GENUS	SPECIES	AUTHORITY	RECON Nr	RECON Name
01 0010020	CARABIDAE	Cicindela	sylvatica	L.	273	Cicindela sylvatica L.
01 0010070	CARABIDAE	Cicindela	campestris	L.	274	Cicindela campestris
01 0020010	CARABIDAE	Calosoma	inquisitor	(L.)	271	CALOSOMA INQUISITOR
01 0020050	CARABIDAE	Calosoma	reticulatum	(F.)	270	Calosoma reticulatum
▶ 01 0040100	CARABIDAE	Carabus	problematicus	Hbst.	283	CARABUS PROBLEMATICUS (ammended distribution)
01 0040120	CARABIDAE	Carabus	granulatus	L.	286	CARABUS GRANULATUS L.
01 0040140	CARABIDAE	Carabus	clathratus	L.	288	Carabus clathratus L.
01 0040172	CARABIDAE	Carabus	convexus	F.	287	CARABUS CONVEXUS F (=TOMOCARABUS)
01 0040180	CARABIDAE	Carabus	nifens	L.	284	CARABUS NITENS L.
01 0040341	CARABIDAE	Carabus	mæander	Fisch.	285	Carabus mæander
01 0050010	CARABIDAE	Cychrus	caraboides	(L.)	253	Cychrus caraboides

Figure 5.3. Screenshot: Comparison of BugsCEP taxa with names of taxa as stored in the original MCRBirm software. The ‘RECON Nr’ field shows the numbers used as input strings in the original RECON(struction) component of the MCRBirm software..

5.1.2.2 Calculations and overlaps

Whilst the earlier software did not calculate temperature values for assemblages where the area of maximum overlap contained < 90 % of the taxa in the sample, it provided the option of sub-setting these faunas on the basis of their warm or cold tolerance (see 5.2.5). BugsMCR currently does not provide warm/cold component estimations, but it does provide the facility to calculate temperatures for any degree of maximum overlap, thus allowing the user to decide on the acceptable minimum percentage. The percentage overlap is given in the ‘Overlap’ column in the results file (see Figure 5.4). It could be postulated that the lower the percentage overlap, the greater the probability that the sample represents a period of time in which there was a degree of climate change greater than the tolerance levels of the individual species found. This is a function of the resolution of the sample, and the rate of climatic change, and not easy to resolve without entering into circular reasoning (see 5.2.5). The broad application of an arbitrary 90 % minimum overlap often prevents reconstruction from large assemblages or samples, where the fauna in its entirety represents the most probable extremes of the temperature that occurred during the terminal depositional lifespan of the sample. Note that the MCR values *do not* represent the average temperature for the period, due to the fact that MCR works on thermal limits rather than means. It is possible that, since MCR calculates the extreme limits of possible temperatures, i.e. the widest ranges, from the processed fauna, the presentation of any reconstructed values are valid as long as the percentage overlap is given. Exceptions may apply in situations where there is evidence, such as differential preservation of fragments or sedimentological indications, of secondary deposition, where the assemblage may represent more than one geographically or chronologically disparate fauna.

For each cell in the 1°C climate space map for a sample, RECON calculated envelope overlaps to the nearest 10 %, as shown in the overlap matrix in Figure 5.4b. BugsMCR theoretically calculates overlaps with up to 28 decimals precision, and can thus more accurately define the area of maximum overlap, when compared to the 90 % or 100 % limits used in RECON. This is sometimes reflected in small differences in the reconstructed temperatures produced by each program – BugsMCR producing a narrower set of ranges than RECON (Figure 5.4). Whether or not this is to be called an increase in accuracy will depend on the user’s acknowledgement of the usefulness of 90 % of the assemblage as an acceptable proportion of the assemblage for reconstruction, or whether a narrower range defined by more taxa is preferable. It may be so that the reliability of a reconstruction is more dependent on the actual species used than the number of them, and more work is clearly needed in assessing the relative reliability of assemblages. A possible method, using jackknifing variants, is suggested below (5.2.6 and 5.2.6.1).

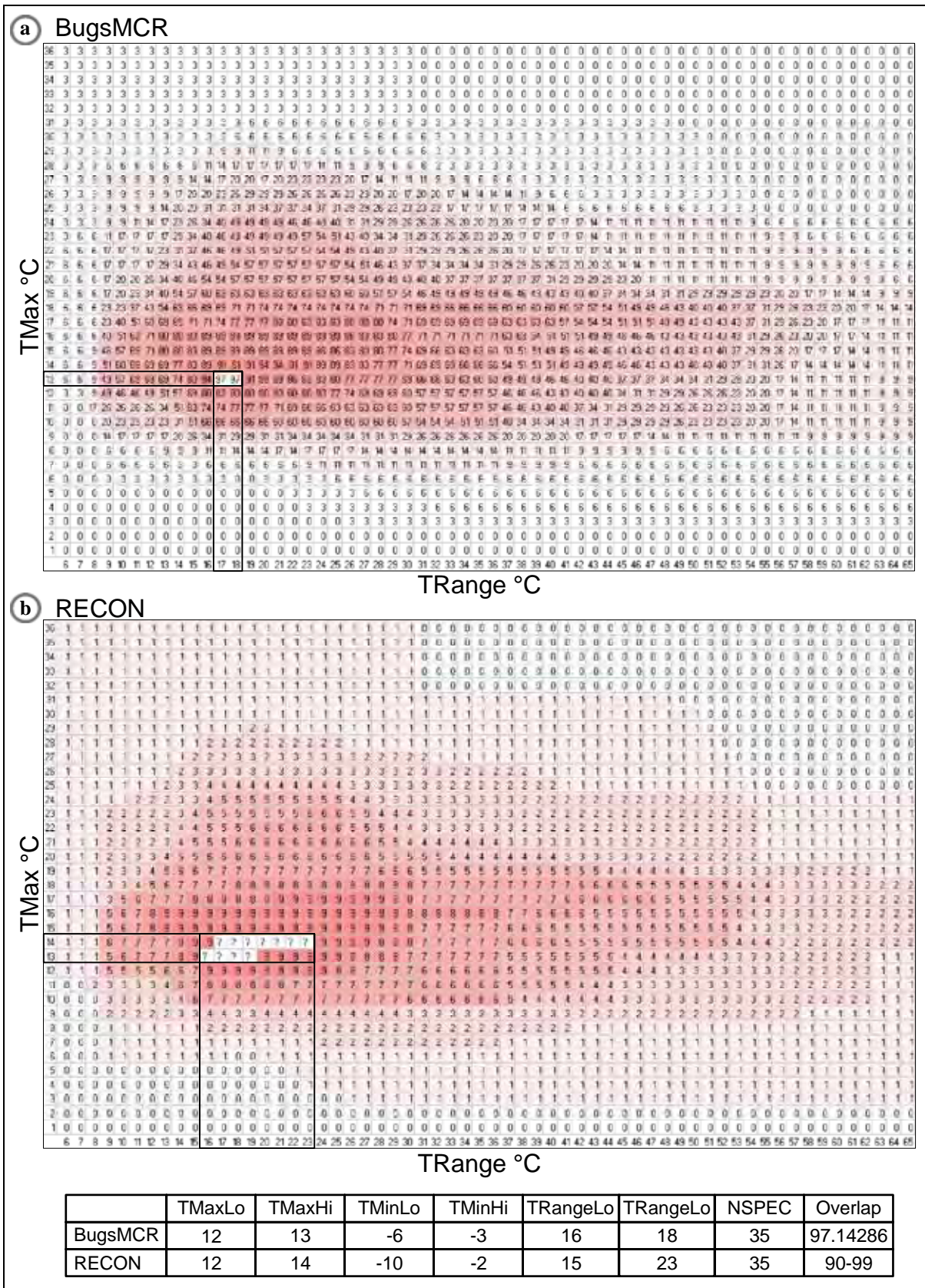


Figure 5.4. Comparison of exported climate space maps for Saint Bees (Coope & Joachim, 1980) sample s50 as processed by (a) BugsMCR and (b) RECON. Note the difference in reconstructed temperatures caused by the definition of maximum overlap cells. Both maps have been shaded, had scales added and the area of maximum overlap highlighted. The BugsMCR output has been rounded to the nearest percent for clarity, and the RECON values are in percent/10 as output by the program. The scales indicate the upper boundary of climate cells in degrees Celsius.

5.1.2.3 Graphs

The original Birmingham MCR software suite included a program called ‘Graphs’ which was able to plot a series of thermal ranges (TMax and TMin) against sample dates. Although BugsCEP does graph the results, there was unfortunately not time to duplicate the facility to scale the x-axis by sample date, and samples are simply plotted in order of entry. The graphing motor in MS Excel is not designed for graphs too far removed from its standard types, and improved scaling would require some trickery in terms of ‘drawing’ with data. A prototype has been created, and this feature will be added at a later date. For the moment users requiring a scaled time axis must import the data into another graphing package, or use graphics software to scale the graphs. In addition, BugsMCR does not attempt to unify the temperature scales of the individual TMax and TMin graphs produced, and this must also be undertaken manually.

5.1.2.4 Advanced MCR

A button in the ‘Tools’ section of the BugsMCR interface give access to the experimental ‘Advanced MCR’ interface (Figure 5.5). This is not fully tested, and the meaning of the statistics that it produces is not yet fully understood^{vi}. The two main features of the advanced interface are the ability to see sample summary data and select samples before running calculations, and the ability to jackknife MCR calculations and produce a number of related statistics. The potential uses for jackknifed MCR calculations are discussed in sections 5.2.6 and 6.7. Multiple removal, or ‘delete-d’ jackknifing can also be performed, and the options are briefly discussed in section 5.2.6.1. The intention is to add more features to this interface with time.

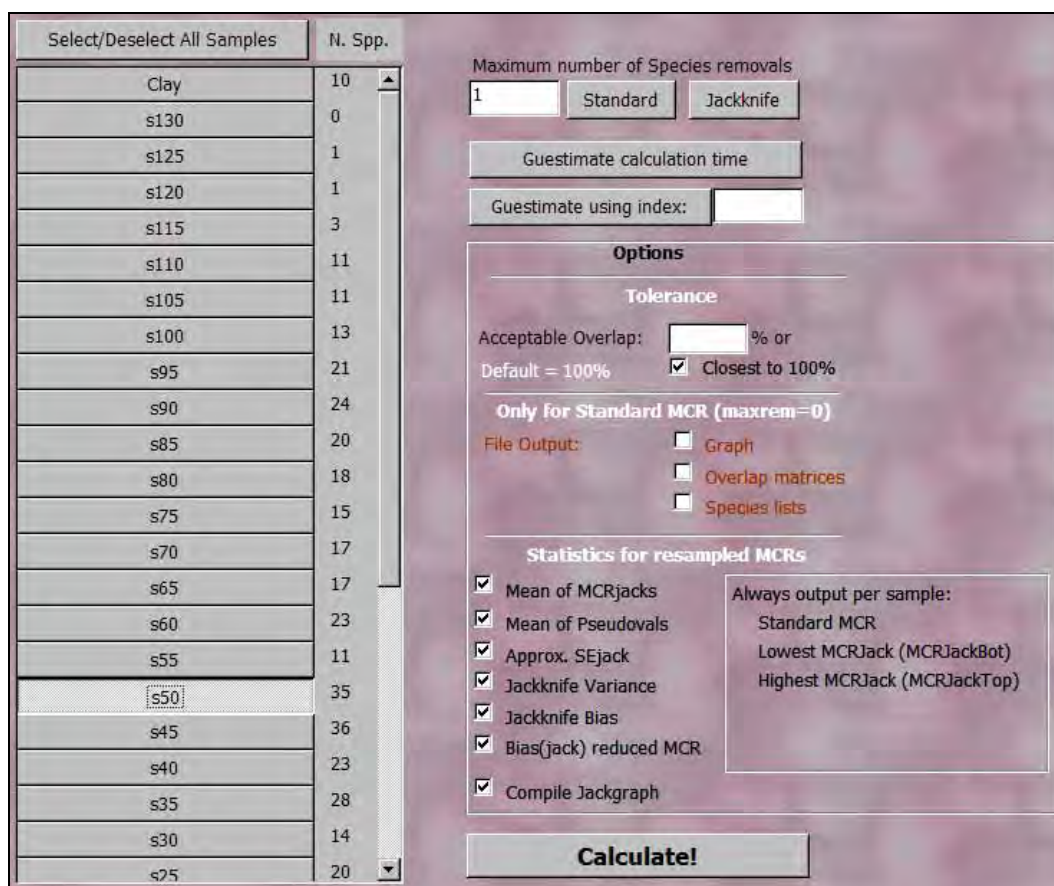


Figure 5.5. Screenshot: Advanced MCR interface, showing the sample selection panel on the left and numerous options on the right.

^{vi} The interface is password protected at the time of publication, and will be made available on its completion.

5.1.3 Predicting potential changes in geographical range

The ‘Predictions’ module of BugsMCR is a simple query engine for retrieving lists of taxa that comply with specific thermal limits or values. It is similar in purpose to the ‘Pest’ program described by Perry (1986), although this was not available for comparison. At the time of writing, the predictions module uses only the extreme limits of the thermal variables, i.e. TMaxHigh, TMaxLow etc, rather than the actual envelopes. This is by design, and allows the system to be slightly optimistic in the species lists it retrieves, although with the loss of some accuracy. It also partly compensates for the poor quality of a number of the species thermal envelopes in the MCR dataset. An alternative, using the actual thermal envelopes, will be implemented at a later date, as this could potentially be a more powerful method.

As Perry (1986) suggests, the MCR method can be inverted, and the thermal tolerance of species used to predict their potential geographical distribution under ideal conditions^{vii}. Perry (1986) used the Pest program to test the viability of beetle distribution by ice rafting, and investigated the percentage of carabid populations found on North Atlantic islands in proportion to the potential richness under different climate scenarios. He showed a clear distance decay model in line with Scandinavian and northern British origins for the faunas of Shetland, Faeroe, Iceland and Greenland. Vickers (2006) has used the BugsMCR predictions module to discuss tabula-rasa/refugia theory in the North Atlantic, and there is considerable potential for experimentation in many research areas from historical biogeography to the prediction of predator or pest species distributions with global warming (see also Vickers & Buckland, *in prep.*).

The predictions pop-up window, accessible by clicking the [Predictions] button on the BugsMCR interface, gives the user access to the thermal limits of all the MCR species, and Figure 5.6 shows those for the stenothermic carabid *Diacheila arctica* (Gyll.). The current distribution of *D. arctica* is restricted to the far north of Scandinavia, North West Russian, and possibly into Siberia, although there is little collection data currently available from Siberia. The beetle is characterised by an extremely narrow tolerance for summer temperatures (TMax), and is known fossil from a number of Late Glacial and Younger Dryas sites in the British Isles, often being cited in evidence for the cold, continental climate of that time (e.g. Ashworth, 2001).

By copying the thermal range data for *D. arctica* into the ‘Ranges’ panel, they become available as search criteria, and the user can simply click the [Find Species that are...] button to see which other species can survive within these limits. As it happens, none can survive within the exact limits of *D. arctica*, which gives us scope for expanding the ranges, or simulating changes in temperature. Any range value may be either adjusted, or omitted, but if specific TMax and TMin values are to be used, then both must be included. TRange, although a useful indication of continentality, and the second most important variable for defining the thermal limits of the beetles in the MCR dataset, is more difficult to visualize than summer and winter temperatures. If the TRange values for *D. arctica* are deleted, and the button clicked again, it opens up the possibility for more, or less, continental species to be found, as can be seen in Table 5.2.

^{vii} A similar, although more complex approach has also been applied for assessing the impact of climate change on plants (Dockerty & Lovett, 2003).

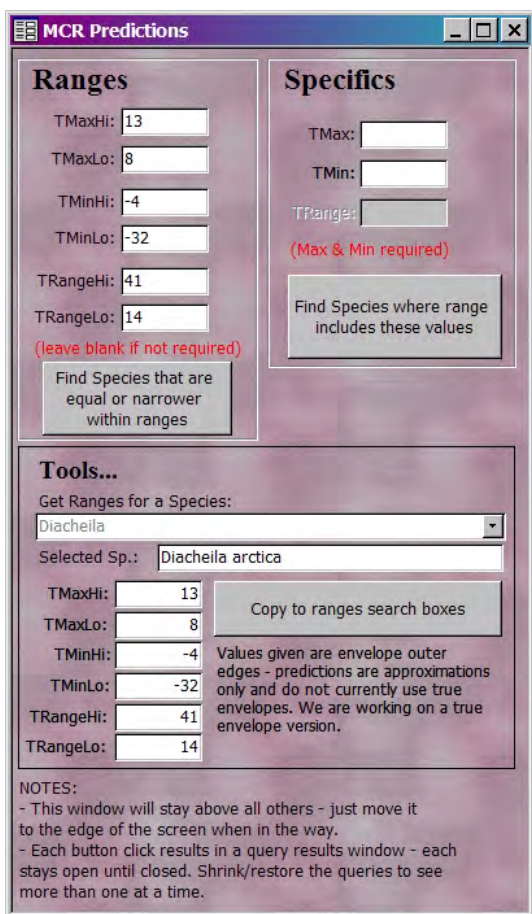


Figure 5.6. Screenshot: The BugsMCR Predictions interface with the thermal envelope limits for *Diacheila arctica* (Gyll.) selected and copied to the ‘Ranges’ panel.

Pterostichus kokeilii Miller (Table 5.2) is slightly more cold stenothermic than *D. arctica*, but does not appear to be able to survive as cold winters. The ranges of these two species do not overlap in geographical space, *P. kokeilii* being restricted to central and eastern European Alpine areas (Koch, 1989; Mani, 1968), but this result shows that they could theoretically exist in the same climate. This is, of course, highly simplified reasoning, and there are a large number of factors that need to be considered when investigating the potential ranges of species during different climatic conditions. Species competition is also a factor to consider when experimenting with ranges. Many of the MCR species are low density predators, and the arrival of a new species due to climate change could upset the predator-prey relationships that exist in an area. In this case, however, *P. kokeilii* is classed as a *Meadowland* species in BugsCEP, whereas *D. arctica* is classed as a *Heathland & moorland* species, and it is perhaps unlikely that they would compete for similar habitats or food sources.

Table 5.2. Species list return from Predictions module when searching for species with TMin and TMax spans equal to or narrower than *Diacheila arctica* (Gyll.).

CODE	Taxon	TMaxLo	TMaxHi	TMinLo	TMinHi	TRangeLo	TRangeHi
01.0111010	<i>Diacheila arctica</i>	8	13	-32	-4	14	41
01.0510410	<i>Pterostichus kokeilii</i>	8	11	-28	-4	13	36

Simulating temperature changes through the adjustment of range extremes is a difficult business, and it is perhaps easier to deal with single temperature values such as mean values for TMax and TMin. The 1961-1990 average temperature for the far northern tip of Sweden is around 10°C for July, and -14°C for January (SMHI, 2005; Naimakka weather station, 68.7° N, 21.5° E). If these values are

entered into the ‘Specifics’ TMax and TMin boxes, and the [Find Species where range includes these values] button pressed, a list of 192 species is retrieved. This surprisingly large list is retrieved due to the fact that all species whose ranges overlap the entered values are found, including the cold stenotherms discussed above, and more eurythermal species such as *Calathus melanocephalus* (L.), which can survive summer temperature from 8°C to 36°C. A simple 2°C global warming scenario can be simulated by raising the TMax and TMin values to 12°C and –12°C respectively, and then running the query again. This retrieves a list of 246 species, more than half the MCR dataset, which could theoretically survive at the new temperatures. It would appear that a uniform two degree rise in the mean summer and winter temperatures would allow more MCR species to survive in northern Sweden. The means, however, give a very limited picture of the climate, as the monthly mean temperatures will have varied from year to year around these values.

For the sake of experiment, this variation can be simulated by setting hypothetical limits to the range of temperatures expected. Focussing on the summer temperature, which is more accurately recorded in the species envelopes, one can produce an alternative list of species that could inhabit the far northern tip of Sweden. Clearing all range boxes and entering 14°C and 7°C for TMaxHi and TMaxLo returns a list of 12 species (Table 5.3). Subsequently raising TMaxHi and TMaxLo by two degrees, to 16°C and 9°C, actually reduces the number of species to 9 (Table 5.4), illustrating the important point that this function retrieves *only* those species that have ranges narrower or equal to the values entered. Rather than just telling us which species could potentially colonise the area with a rise in temperature, this experiment has told us which species are at risk of local extinction (although it must be remembered that not all of those listed will actually live at the location – the MCR software has no geographical awareness). In addition, the experiment retrieves a list of several, slightly more warm tolerant, stenothermic species which could theoretically move in. These would, according to their thermal envelopes, not have been able to survive the lower extremes of the 7-14°C TMax summers. The advantage of this method over the use of means, is that the maximum survival limits of species are taken into consideration, and thus the retrieved list are considerably shorter.

Additional query functions, and the refinement of the prediction module to include envelope, rather than range, based querying would allow these lines of enquiry to be explored in more detail, and build up a fuller picture of potential distribution changes. At the moment, prediction envelopes can only be seen by importing the query results (e.g. Table 5.3) into a BugsCEP site and running MCR on the list.

Table 5.3. List of species, and their thermal tolerance extremes, which have TMax ranges equal to or narrower than 7-14°C. Species that would not be able to cope with an increase in summer temperature ranges of 2°C are in **bold**.

CODE	Taxon	TMaxLo	TMaxHi	TMinLo	TMinHi	TRangeLo	TRangeHi
01.0111010	Diacheila arctica	8	13	-32	-4	14	41
01.0111020	Diacheila polita	7	12	-38	-5	15	49
01.0120007	Elaphrus lapponicus	9	13	-36	3	10	47
01.0290045	Bembidion lapponicum	8	14	-37	-5	19	49
01.0290141	Bembidion hyperboreaorum	8	13	-39	-7	20	47
01.0510410	Pterostichus kokeilii	8	11	-28	-4	13	36
01.0510578	Pterostichus middendorffi	7	11	-39	-8	19	46
01.0510580	Pterostichus vermiculosus	7	11	-40	-17	28	47
01.0620071	Agonum chalconatum	10	14	-25	-6	20	35
01.0620285	Agonum consimile	9	14	-27	-2	14	36
23.0260081	Olophrum boreale	8	14	-21	0	14	29
23.0300043	Acidota quadrata	9	13	-30	-2	15	39

Table 5.4. List of species, and their thermal tolerance extremes, which have TMax ranges equal to or narrower than 9-16°C. Stenothermic species that would not have been able to survive a 2°C lower TMax value are in **bold**.

CODE	Taxon	TMaxLo	TMaxHi	TMinLo	TMinHi	TRangeLo	TRangeHi
01.0120007	<i>Elaphrus lapponicus</i>	9	13	-36	3	10	47
01.0290240	<i>Bembidion virens</i>	9	16	-27	5	8	36
01.0290541	<i>Bembidion petrosus</i>	10	16	-37	1	13	49
01.0620071	<i>Agonum chalconatum</i>	10	14	-25	-6	20	35
01.0620285	<i>Agonum consimile</i>	9	14	-27	-2	14	36
23.0240010	<i>Deliphrum tectum</i>	9	16	-21	5	10	30
23.0260121	<i>Olophrum rotundicolle</i>	9	15	-21	2	13	30
23.0300043	<i>Acidota quadrata</i>	9	13	-30	-2	15	39
23.0350060	<i>Anthophagus bicornis</i>	9	15	-16	1	14	25

Although the majority of species in the MCR list have been chosen for their lack of ties to specific vegetation types, it is quite probable that a number of them still have other specific habitat needs. The ecology of many species is poorly understood, especially at their geographical or climatic limits, and it has previously been argued that regional variation in the microhabitat preference of carabids reduces the usefulness of MCR as a climate reconstruction method (Andersen, 1993; 1996). Coope & Lemdahl (1996) argue that the microclimate (as described by Andersen, 1993) exists as a subcomponent of the macroclimate used when constructing the species thermal envelopes, which are built from the entire geographical range of the species, and these variations are thus catered for in the method. Even so, care must be taken to look at the geographical possibility for species migrations, past or future, when predicting ranges from climate data. The problems of habitats and geographical ranges are discussed further in section 6.7.

In the examples given above, the influence of the dataset scope on retrieved species lists is potentially large. The fact that the MCR dataset has a large Scandinavian component makes this less of a problem in this case, but should the same experiments be carried out for Central or Eastern Europe, the results would probably be unreliable in that a large number of important species would be missing. There is, as is mentioned at several places in this chapter, a great need for the enhancement and expansion of the MCR dataset. In addition, the combination of the prediction module with a GIS system would greatly increase the potential for investigating the past, present and future biogeography of beetles, perhaps by including their present geographical distributions in the prediction algorithms. The addition of phytophagous species, for example, could make the system interesting for forestry and agriculture in the prediction of changing pest patterns due to global warming. It is possible that additional variables would have to be added to complement TMax and TMin, as these may not be the limiting factors for all species (Parmesan *et al.*, 2005).

5.2 MCR – Problems and Possibilities

5.2.1 Requirements for quantitative reconstructions

Birks (2003) lists nine major requirements for quantitative reconstructions, two of which the MCR method could be considered as failing to meet. Firstly, the method does not “...model the non-linear relationships between modern taxa and their environment and take account of the numerical properties of the biological data.”. There are a number of reasons for this, the primary one being that MCR is a presence/absence method by definition, and numbers of individuals are, mathematically, uninteresting. A single find of a species has as much weight in the final reconstruction as any number of individuals, and this allows MCR to reconstruct temperatures from very poor faunas. There is, of course, a risk that the identifications are incorrect, or that the sample does not represent the environment that is to be reconstructed due to redeposition or contamination, for example. These are, however, taphonomic concerns, and the domain of the individual doing the identification and processing the samples, and

these are factors that cannot easily be quantified in the reconstruction methodology without resorting to guesswork. This is true for any quantitative reconstruction method, irrespective of proxy data source. Palaeoentomology has the advantage, however, that the taxonomic resolution of identifications is most often far greater than that possible for a number of other proxies.

A number of authors have attempted to either build population density functions around the original collection data, or apply general population response models to the thermal envelopes in an effort to access the information potentially contained in abundance data (e.g. Huppert & Solow, 2004). The former method is ambitious, and requires not only a considerable improvement in the amount of beetle collection data, but also the application of a number of assumptions upon that data in order to smooth it into a probability density function for the species. The quality of the underlying distribution data are variable, and spatial abundance data are rare, the majority of beetle distribution data being point finds. In addition, pitfall trap data are often biased, being the results of habitat specific surveys, and often must be statistically treated before they can be used to estimate reliably the population densities of species (Krebs, 1999; Southwood & Henderson, 2000; Brown & Lomolino, 1998). It is possible to combine species data collected by multiple methods (e.g. Graham & Hijmans, 2006), but whether abundance data simulated from this would be any more realistic than general models is difficult to assess.

Theoretically, the population density of any species will vary within its climatic range, tending to decay towards the edges of survivability. A unimodal model is often chosen to approximate this, with a central peak spread over the theoretical optimum levels of each variable. Although the models work well with a number of statistical treatments, and allow for the estimation of errors, there is limited empirical evidence supporting their use to describe abundance or population density as a response to temperature. They should therefore be treated with as much care as presence/absence models, or data based simulations, especially for broad applications over wide geographical areas. Unimodality is a reasonable assumption, but applications often ignore the potential for skewed distributions or thresholds (such as freezing) in response to variables. There has been a large amount of recent work assessing the various attempts to model the limits of species ranges in terms of environmental variables (cf. Parmesan *et al.*, 2005). A comparison of a number of models, along with an alternative to MCR for New Zealand, is presented by Marra *et al.* (2004).

At the other end of the reconstruction process is the fossil beetle assemblage which should represent the palaeoclimate in question. Due to the problems of taphonomic loss and sample representativeness, it is difficult to know as to whether the abundance patterns in the sample accurately represent the species populations that were alive at the time of deposition. As a result of this, it is equally difficult to know whether the abundance data are relevant to the particular model being applied. Again, this is an issue that permeates all proxy data sources in all forms of reconstruction, and must be considered during the application of models and the interpretation of their results. One should, of course remember that “all models are wrong, some are useful” (Box, 1987: p424), and that the above concepts may still provide useful approximations to reality that allow us to model the past more accurately. More work is certainly necessary in this area, and cooperation between ecologists, (palaeo-)entomologists and biogeographers is essential.

The second of Birks' (2003) criteria not met by the MCR method in its coleopteran form is that of the reliable estimation of sample specific errors. Quaternary science is accustomed to the presentation of mean and standard deviation values for temperature reconstructions, and Atkinson *et al.* (1987) attempted to provide this from the MCR method by constructing a linear regression based correction equation. This method has problems (see section 5.2.3), and there is currently no other method used for providing a measure of the reliability of an MCR reconstruction. The use of resampling statistics may provide at least part of a solution, allowing the internal reliability of each set of species to be tested, as discussed below (5.2.6).

5.2.2 MCR and urban deposits

When reconstructing palaeotemperatures it is very important to be aware of what the assemblage in the samples represents. Climate reconstructions have sometimes been undertaken from archaeological urban deposits (e.g. Hellqvist & Lemdahl, 1996), and although there is evidence that some species in these deposits may reflect background environments (Kenward, 1985), the interpretation may be problematic. The effect of urban areas is not only to warm the ambient temperature, and often reduce the variability of annual temperatures, in a dome like ‘urban heat island’ (Camilloni & Barros, 1997), but also create artificially warm environments which allow species to survive in regions outside of their natural geographical range. They often provide a variety of indoor and ground level microhabitats which can support a wider fauna than the surrounding natural environment. These microhabitats often have raised temperatures, due to human activities such as heating, covering and waste disposal, which allow species that would not naturally be able to survive in the area to maintain populations. This is particularly evident from northern Europe, where work on Norse and medieval farm sites has provided considerable evidence for the artificially warmed interior of structures supporting species with more southerly origins (Buckland *et al.*, 1996). If the beetles that occupy these environments are also used to reconstruct temperatures using MCR, then they could lead to warmer reconstructions than the ambient temperature outside of the settlement. By using MCR on an urban fauna it is the temperature of these additional microhabitats that are most likely being reconstructed, and not the regional temperature, or even the urban heat island, as some would assume. The variable nature of such microhabitat creation, added to the possibilities for a context to capture individuals from them, makes it extremely difficult to interpret such an MCR signal. To try to deconstruct the fauna then into local and background components, from which different reconstructions can be made, is a dangerous game which can easily lead to a reliance on circular reasoning.

Although more work is necessary on urban deposits, it should be undertaken cautiously with respect to temperature reconstruction, which should only be undertaken along with a thorough investigation into the habitats represented by the species used for the thermal reconstruction, in addition to the environments represented by the rest of the fauna. There are many potentially interesting research angles related to archaeological deposits and biogeography, including dispersal issues and the persistence of species upon the abandonment of sites.

5.2.3 Problems with MCR correction/calibration

The MCR method’s ability to reconstruct palaeotemperatures has been successfully validated by the reconstruction of independent modern faunas (Atkinson *et al.*, 1987). The validation method has also been suggested for correcting MCR temperatures to compensate for the systematic bias exhibited by MCR values when tested on an independent modern dataset. The calibration technique uses linear regression of median values to derive formulae for correcting TMax for its tendency to reconstruct the summers warm climates as too cold, and TMin for its tendency to underestimate the severity of cold climates. It is also often cited as the most probable value of the MCR, with the corrected values providing “...unbiased estimates of the most probable palaeoclimate within the MCR...” (Atkinson *et al.*, 1987: p588; but see also Walkling & Coope, 1996 and Coope *et al.*, 1998, where the equation is refined). Unfortunately this is not true. Due to the binary nature of the species thermal envelopes (Figure 5.2), each and every grid square is equally probable to represent a climate in which the species can survive. A consequence of this is that the area of maximum overlap for any group of species also represents a uniform probability surface (Figure 5.7) – there is nothing in the method that allows for one grid cell of this area to be more probable than any other, including the median value. In addition, the median, defined as the middle value between the extremes of an envelope, may not necessarily be the most appropriate measure of centre for either a species envelope or an MCR, due to the non-symmetrical nature of the matrices, and the variable nature of their source data.

A further problem with the calibration method is the assertion of precision (c. $\pm 2^{\circ}\text{C}$ for TMax and c. $\pm 5^{\circ}\text{C}$ for TMin), which, although correctly derived for the regression equation, it is doubtful as to whether these values can be reliably transferred to other reconstructions. Standard deviation is the

RMS (root mean square) deviation of values from an arithmetic mean, in this case derived from the linear regression comparing reconstructed medians with independent modern data. The errors, then, are for the calibration attempt, and not the actual reconstructed medians themselves. In addition, the quality of the species envelope data is variable, and hence assemblages will vary in their internal reliability. The quantification of the internal reliability of assemblages is discussed below (5.2.6), but such errors are difficult to assess, and an evaluation of individual envelope reliability (as suggested in Perry, 1986) should be used to better quantify errors.

An alternative to calibrating the median would be to calibrate the range limits (TMaxLo, TMaxHi, etc.), in the same way as described above. Unlike the median, range limits are significant values, representing the outer bounds of the uniform area of greatest overlap. Testing of these values against an independent modern dataset, and the derivation of standard errors through regression could help identify the accuracy of the range boundaries. This does not, however, address the problem of quantifying the reliability of the reference data, which is a problem common to any quantitative reconstruction method, and rarely evaluated. It also represents a loss of information in that the species envelopes are not rectangular, and thus the range limits only have a limited ability to describe them.

The question still remains as to whether there is actually any real reason for trying to estimate the most probable value within the range. Palaeoecological samples, especially those taken for beetles, almost always encompass a variable time depth, from a moment in time in some archaeological deposits (the ‘Pompeii effect’ (Schiffer, 1987)), to several hundred, if not several thousand years. The climate within any extended time span will most likely have varied, and in many cases an indication of the range of temperatures represented by the assemblage will be more useful than an mean value. Calculation of the mean value represents a significant simplification of the available information when compared to the range limits. The demands for mean and standard deviation, not least amongst the Quaternary community, and the desire to present more and more ‘accurate’ reconstructions in all measurements may be largely to blame for the aversion to specifying thermal limits.

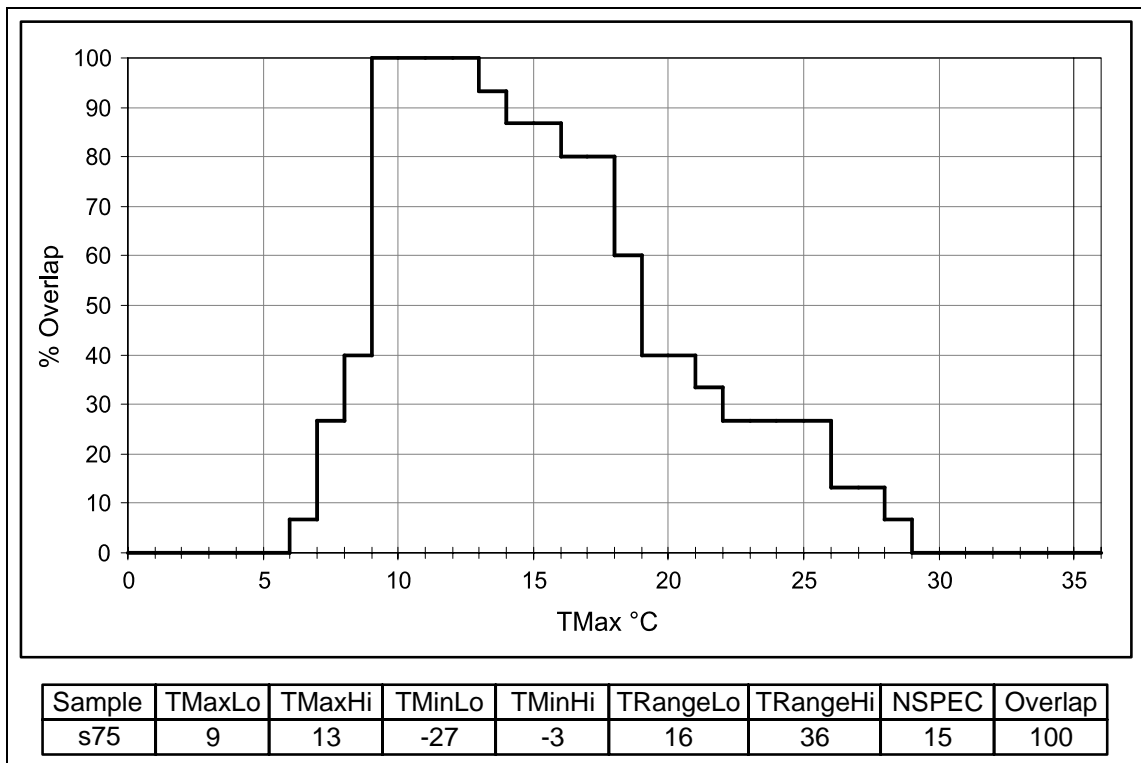


Figure 5.7. Cross section of a sample climate space map, along the TRange = 21°C line, showing % of species in each TMax 1°C climate cell. Note the square nature of the curve, and the plateau like nature of the 100 % area, indicating the equal probability of any cell in this area.

5.2.4 Ubiquity analysis

Bray *et al.* (2006) suggest the use of ubiquity analysis to refine the MCR estimates, although they do not present an actual method for obtaining temperature values. Bray *et al.* map the point occurrence data of the original MCR reference dataset into frequency of occurrence in gridded European climate cells. This grid is then used to derive a ‘ubiquity score’ by standardizing the occurrence of each species in a particular climate cell against the frequency of occurrence of the climate cell in Europe. The ‘ubiquity score’ is essentially a measure of the probability of occurrence of a particular species in a particular climate cell. Unfortunately, the method compounds the problems inherent in the irregularity of insect collection data, and produces a frequency of thermal occurrence map which is entirely tied to the accuracy of the base data. In other words the system does not show “...how much of a particular sector of climate space ... a species actually occupies” (Bray *et al.*, 2006), but the relative frequency of collection of a species in a particular climate sector. Whilst the method undoubtedly has research potential, utilizing redundant information in the original dataset, it does not seem to provide any refinements to the accuracy of reconstructions as yet. In common with the original MCR method, it is also severely restricted to the range of climate space cells available in the original dataset, and its ability to extrapolate beyond appears limited without addition to the MCR base data. Criticisms aside, ubiquity analysis appears to present one potential method for constructing a data based probability density function for approximating the thermal envelopes of species in more dimensions. The problems in the reliability and scope of the base, thermal envelope, dataset remain, and must be resolved before any method, no matter how it manipulates the data, can be expected to refine the accuracy of reconstructions.

5.2.5 Calculation of relative warm/cold components

The RECON software used the overlap of species envelopes with two somewhat arbitrarily selected grid cells to identify warm and cold faunal components (Table 5.5) in assemblages that provided too little overlap for MCR calculation (Perry, 1986). This could be used to help provide information on thermal anomalies with shorter durations than the resolution of the samples, separate out the extreme components of a gradient, or identify other reasons for the non-overlap situation. These cells had “...no intrinsic merit” (Perry, 1986), and gave a somewhat abstract division of assemblages, and, although useful, limited the investigative power of the function.

Table 5.5. Warm/Cold reference cells used in RECON.
Reproduced from Perry (1986:131).

	TMax	TRange
Warm reference	17 - 18°C	19 - 20°C
Cold reference	8 - 9°C	30 - 31°C

These reference values are most certainly not adequate for dividing all assemblages, and a system has been devised, through the work behind this thesis, to use the centre of gravities (COG)^{viii} of the individual species envelopes to define the warm/cold reference cells on a sample by sample basis. BugsMCR calculates a sample specific COG using all the species in a sample, and then compares the relative warm/cold position of each species envelope COG to this. As there are a number of options for calculating a COG, this allows for some flexibility and experimentation in the resolution of warm/cold faunal components. It also provides the possibility for a more context specific breakdown of non-overlap assemblages, which may help resolve the reasons behind the diversity of the fauna. Most importantly, it allows for the separation of relative measures of warm/cold, e.g. a relatively cold and relatively warm component within a generally warm phase. This could possibly be especially useful in helping to resolve short duration events, such as the 8 200 cal. yr BP event.

^{viii} Note that this should not be confused with the envelope means or medians, which are specific measures of central tendency, and can be used as methods in the calculation of a COG.

Care must be taken, however, as was equally the case when using RECON, not to give these experimental reconstructions the same worth as those calculated from full assemblages. The COG of the unsymmetrical binary species envelopes is in itself somewhat arbitrary, and although it has useful scientific value, there is no statistical foundation behind it at present. This module is currently incomplete, but will be made available in the near future as an experimental tool.

5.2.6 The use of jackknifing to investigate or enhance the reliability of results

The lack of standard error calculation in the MCR method has led to some criticism of its use, and MCR is sometimes overlooked as a viable quantitative method as a result of this (e.g. Birks, 2003). Apart from the successful testing of the method against independent modern data (Atkinson *et al.*, 1987), from which the regression based correction technique described above (section 5.2.3) is derived, the MCR method, as applied to beetles, has no internal verification capacity. Although Witte *et al.* (1998) use bootstrapping and Monte Carlo simulations to test the reliability of thermal gradients derived through MCR, no attempt has been made to test the reliability of the MCR reconstructions themselves using resampling.

The resampling method of jackknifing, or leave-one-out analysis, has been used in a variety of fields to provide a measure of internal reliability of data sets (e.g. Manly, 1997). Improvements in computer performance have been instrumental in the increasing use of resampling methods, which are by their nature computationally intensive. The system described below has been developed, and tested, but is not in a state sufficiently advanced or understood to be comprehensively utilized. There are still many questions that need to be answered on the validity and interpretation of the results, and these will be presented in a forthcoming publication. The intention is to construct software that is able to provide an indication of the reliability of reconstructed sample temperature range limits (TMaxHi, TMaxLo, etc.). A worked example is provided in Chapter 6, and a more detailed examination of a single sample from the Saint Bees' site (Coope & Joachim, 1980) is given below.

Resampling techniques are a set of conveniently simple methods for simulating larger datasets using only the data available, and in many cases providing confidence intervals for the results obtained. A number of resampling methods are available, but few are useful on presence/absence data. Jackknifing works, as do most resampling methods, by duplicating the current dataset with a minor variation, and then running the original calculations on the new dataset. In the case of jackknifing, the variation is caused by the removal of one data item, or in our case taxon, and replacing it before removing the next. The process is simple, as follows:

1. Calculate MCR on the full dataset
2. Remove first taxon, and calculate MCR on the reduced dataset (MCR_{J1})
3. Replace the first taxon, remove the second, and calculate MCR on the new reduced dataset (MCR_{J2})
4. Repeat the process for all taxa, so that a set of MCR_{Ji} results is created
5. Calculate statistics and limits from the results

The computer implementation is equally simple, in the form of a couple of nested loops and counters to provide the taxon removal sequence, and simply calling the existing MCR calculation module on each inner loop. The manipulation of the output data, which is geometrically proportional to the number of taxa in each sample, however, is not so simple. After initial tests using tables and queries, a VBA solution was arrived at, due to the former leading to enormous expansion and instability in the database. The routine performs calculations entirely in arrays in memory, and dumps results into a prepared MS Excel file. Statistics are similarly calculated from the results array, stored in a new array, and then dumped to MS Excel. A number of worksheets are used to organize the output so that it is more manageable, and the user can select which of these to include (see Figure 5.5). Results are also output in a form which may easily be graphed using MS Excel's built in 'Stock Market Open-High-Low-Close' chart (Table 5.6), although some additional manipulation may be necessary to produce ideal graphics (Figure 5.8).

The results of each jackknifed calculation are exported, to allow users to examine the implications of each and every species removal. In the example provided below (Table 5.7), it can be seen that only the removal of two species (in *italics*) lead to a change in the MCR reconstruction. The example represents a stable reconstruction, where only about 5.7 % of the species removals cause a change in the reconstructed temperatures. The two species concerned, *Helophorus sibiricus* (Mots.) and *Boreaphilus henningianus* Sahl., both have arctic distributions that extend into Siberia (Hansen, 1987; Mani, 1968), but they affect the MCR in slightly different ways. The hydrophilid *H. sibiricus* has a poorly defined thermal envelope, but appears to prefer a more continental climate, and its removal expands the TRange lower limit, suggesting a very slightly less continental climate. Its removal also decreases the upper value of TMin by 1°C, suggesting slightly colder winters, although this is probably a consequence of the approximate nature of the TRange extension of the envelope. *B. henningianus* is a relatively cold stenothermic wetland species, the removal of which leads to a slightly warmer TMax, colder TMin and more continental TRange, the last two being a consequence of each other. Closer examination of the MCR output shows that removal of this species highlights a bimodality in the overlap matrix, symptomatic of mixed faunas, which would require further investigation if this experiment was to be followed up. *Calathus erratus* (Sahl.) (in bold in Table 5.7) could be responsible for this, being the only species whose removal leads to 100 % overlap of the remaining species envelopes. It is a relatively warm and eurythermal species when compared to the rest of the fauna.

Where one sample is being examined, the extended limits produced by the jackknifing routine can be plotted on a single graph for convenience (Figure 5.8), or overlain onto the standard TMax and TMin graph where a sequence of samples are examined (see example in Chapter 6). The percentage of species the removal of which causes a change in the reconstructed temperatures, output as *PctSppTDiff* in the results files (Table 5.8), may in itself be a useful indication of the internal reliability of samples. Jackknifing gives the opportunity to produce a number of other statistics from the resampled results, some of which may be used to assess the stability of TValues with species removals.

Table 5.6. Jackknife statistics for sample s50, Saint Bees (Coope & Joachim, 1980), calculated from the results shown in Table 5.7. See Table 5.8 for more details.

Sample	Stat	TMaxLo	TMaxHi	TMinLo	TMinHi	TRangeLo	TRangeHi
s50	MCRStd	12	13	-6	-3	16	18
s50	MCRJackBot	12	13	-8	-3	15	18
s50	MCRJackTop	12	14	-6	-2	16	21
s50	VJack	0	0.94	3.77	0.94	0.94	8.49
s50	BJack	0	0.97	-1.94	0.97	-0.97	2.91
s50	MCRJackMean	12	13.03	-6.06	-2.97	15.97	18.09
s50	PseudoMean	12	12.03	-4.06	-3.97	16.97	15.09
s50	SEJack	0	0.97	1.94	0.97	0.97	2.91
s50	BRedJack	12	12.03	-4.06	-3.97	16.97	15.09

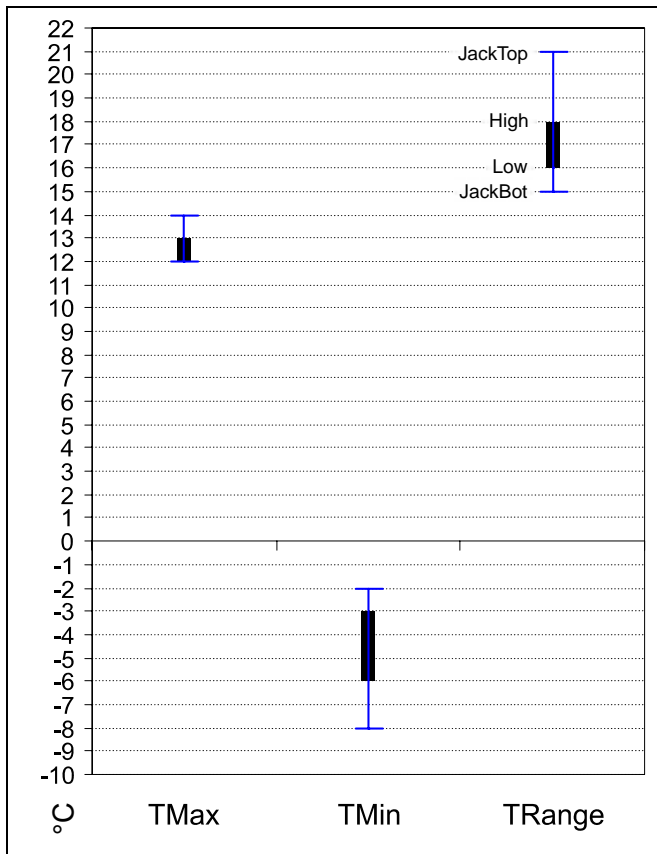


Figure 5.8. Chart showing jackknife limits for all temperature variables for Saint Bees sample s50. The thick bars show the standard MCR results, and the extensions the jackknife extremes. The source data is presented in Table 5.6.

Table 5.7. Jackknife process output for sample s50, Saint Bees site, slightly modified as follows: Species whose removal causes a change in the TValues are in *italics*; species whose removal leads to a 100 % overlap area are in **bold**.

Sample	TMaxLo	TMaxHi	TMinLo	TMinHi	TRangeLo	TRangeHi	NSPEC	Overlap	NSpecRemoved	RemovedCODE1	Species
s50	12	13	-6	-3	16	18	35	97.14286	0		
s50	12	13	-6	-3	16	18	34	97.05882	1	1.00401	Carabus problematicus
s50	12	13	-6	-3	16	18	34	97.05882	1	1.021001	Epaphius secalis
s50	12	13	-6	-3	16	18	34	97.05882	1	1.029092	Bembidion doris
s50	12	13	-6	-3	16	18	34	97.05882	1	1.032001	Patrobus septentrionis
s50	12	13	-6	-3	16	18	34	97.05882	1	1.032002	Patrobus assimilis
s50	12	13	-6	-3	16	18	34	97.05882	1	1.051019	Pterostichus nigrita
s50	12	13	-6	-3	16	18	34	97.05882	1	1.051022	Pterostichus minor
s50	12	13	-6	-3	16	18	34	100	1	1.056002	Calathus erratus
s50	12	13	-6	-3	16	18	34	97.05882	1	1.056006	Calathus melanocephalus
s50	12	13	-6	-3	16	18	34	97.05882	1	1.062028	Agonum fuliginosum
s50	12	13	-6	-3	16	18	34	97.05882	1	1.0620285	Agonum consimile
s50	12	13	-6	-3	16	18	34	97.05882	1	4.007002	Hygrotus inaequalis
s50	12	13	-6	-3	16	18	34	97.05882	1	4.007003	Hygrotus quinquelineatus
s50	12	13	-6	-3	16	18	34	97.05882	1	4.0080015	Hydroporus notabilis
s50	12	13	-6	-3	16	18	34	97.05882	1	4.008005	Hydroporus tristis
s50	12	13	-6	-3	16	18	34	97.05882	1	4.008012	Hydroporus erythrocephalus
s50	12	13	-6	-3	16	18	34	97.05882	1	4.023022	Agabus congener
s50	12	13	-6	-3	16	18	34	97.05882	1	4.0230261	Agabus arcticus
s50	12	13	-6	-3	16	18	34	97.05882	1	4.024005	Ilybius subaeneus
s50	12	13	-6	-3	16	18	34	97.05882	1	4.027001	Colymbetes fuscus
s50	12	13	-6	-3	16	18	34	97.05882	1	5.002003	Gyrinus marinus
<i>s50</i>	<i>12</i>	<i>13</i>	<i>-6</i>	<i>-2</i>	<i>15</i>	<i>18</i>	<i>34</i>	<i>97.05882</i>	<i>1</i>	<i>7.0050065</i>	<i>Helophorus sibiricus</i>
s50	12	13	-6	-3	16	18	34	97.05882	1	9.003019	Cercyon tristis
s50	12	13	-6	-3	16	18	34	97.05882	1	9.008001	Hydrobius fuscipes
s50	12	13	-6	-3	16	18	34	97.05882	1	9.015001	Chaetarthria seminulum
s50	12	13	-6	-3	16	18	34	97.05882	1	23.0120013	Pycnoglypta lurida
s50	12	13	-6	-3	16	18	34	97.05882	1	23.024001	Deliphrum tectum
s50	12	13	-6	-3	16	18	34	97.05882	1	23.026004	Olophrum fuscum
s50	12	13	-6	-3	16	18	34	97.05882	1	23.026006	Olophrum assimile
s50	12	13	-6	-3	16	18	34	97.05882	1	23.0260081	Olophrum boreale
s50	12	13	-6	-3	16	18	34	97.05882	1	23.028001	Eucnecosum brachypterum
s50	12	13	-6	-3	16	18	34	97.05882	1	23.030001	Acidota crenata
s50	12	13	-6	-3	16	18	34	97.05882	1	23.033003	Geodromicus nigrita
<i>s50</i>	<i>12</i>	<i>14</i>	<i>-8</i>	<i>-3</i>	<i>16</i>	<i>21</i>	<i>34</i>	<i>97.05882</i>	<i>1</i>	<i>23.038001</i>	<i>Boreaphilus henningianus</i>
s50	12	13	-6	-3	16	18	34	97.05882	1	23.117017	Tachinus corticinus

Table 5.8. Statistics calculated by the BugsMCR jackknife routine.

Abbreviation	Description
MCRStd	Standard MCR calculation results (TMaxHi, TMaxLo etc.)
MCRJackTop and MCRJackBot	The extreme values created by the jackknifed calculations
MeanOverJ	The mean maximum percentage overlap produced by the jackknifed faunas
PctSppTDiff	The percentage of species whose removal caused a change in any of the temperature variables
Jackknife statistics	For each variable, the following are calculated from the jackknife results: VJack – jackknife variance BJack – jackknife bias MCRJackMean- mean temperature of the jackknife values PseudoMean – used only in calculations SEJack – jackknife standard error BRedJack – Bias reduced jackknife TValue

Preliminary testing on a small number of sites suggests that jackknife standard error (SEJack) is often loosely proportional to $1/N_{\text{Spec}}$, and that PctSppTDiff is generally proportional to $1/\log(N_{\text{Spec}})$, although this may be strongly influenced by the choice of test sites. More work is needed before the accuracy, usefulness or even validity of the other statistics can be discussed, although the system appears to work at least as an informal investigative tools.

5.2.6.1 Multiple removal jackknifing as a reliability index

As an experiment, the jackknifing routine was enhanced so that an increasing number of species could be removed on consecutive passes, a method sometimes known as delete-d, or multiple delete-d jackknifing. In other words, after running a jackknifed MCR with each and every taxon removed individually, the routine was run again with each and every combination of two taxa removed, then three, and so on, until the number of taxa in the sample was reached. This could then be used to plot changes in reconstructed temperature values against numbers of taxa removed, and thus provide an indication of the reconstruction stability inherent in a fauna.

This is an extremely computer intensive routine, but is proving to provide some interesting results. Preliminarily, it can be seen that faunas display threshold numbers beyond which reconstruction quality declines, although it is too early to say if there are any patterns related to the type of faunas involved. The reconstruction of some samples, most likely those that cover climatic gradients, the pattern appears to be more complex, and the data management involved in interpreting these is proving restrictive. New summary measures and automated data manipulation routines will be programmed, and this work will be published separately in the near future.

5.2.6.2 Potential problems with jackknifing and MCR

Whilst the jackknife processes outlined above undoubtedly produce useful results for investigating the implications of the species used in a reconstruction, there is a possibility that some of the statistics are invalid under certain situations. In particular, jackknifed MCRs that produce no thermal values will upset jackknife means, unless they are excluded from the calculations. Jackknife estimators may also not be valid where resampling gives the same values as standard calculations (MCR) (M. Ekström, *pers. comm.*), although this must be investigated more thoroughly. There is, however, evidence that delete-d jackknifing can provide more consistent standard errors than the standard jackknife (Chi & Russell [online], 1999), although as to whether this is true for MCR data is yet to be thoroughly investigated.

In addition, jackknifed values rarely lead to a reduction in the apparent extent of the area of maximum overlap, as defined by the temperature values. This is methodologically logical where simple sample

overlaps are concerned, but the effects of more complex, multimodal overlap scenarios on the statistics require further examination at the individual sample level.

5.3 Conclusions and Future Directions

An advantage of the MCR method is that it calculates on the presence/absence of species, which allows for reconstructions based on very low abundances. In fact, abundance has no effect on reconstructions, although this in itself could be considered as detrimental to the method's ability to provide error estimations for reconstructions. The current state of knowledge on the effects of climate on the relative abundance of all species is, however, too poor for this aspect of the MCR method to be developed further without a considerable amount of work on the primary dataset. There have been recent attempts to extract a probability based abundance surrogate, in the form of the probability of occurrence (presence) of taxa in European gridded climate cells, under the name of ubiquity analysis (Bray *et al.*, 2006). Although this method reveals interesting information on the climate preference of the taxa *as represented in the MCR dataset*, it has a number of problems which currently prevent it from improving on the existing MCR method. The probabilities derived are limited to the scope of the dataset, a problem common to any method that derives secondary calibration data from the primary calibration set^{ix}. The use of resampling methods can provide a measure of reliability by supplementing an independent control dataset with (at least n-1) resampled calculations, and may thus be able derive resampled standard errors.

What is essential at this point is an expansion of the calibration set to include more species, especially continental ones, and a refinement of the envelopes for the existing species. With present day computing power, there is little need for the 1°C cell resolution, and more complex envelopes can be stored and manipulated relatively simply due to improvements in database management and programming tools (although one could question whether a greater resolution would actually lead to a *real* increase in reconstruction accuracy). The collation of this data would, as always, probably be the most time consuming part of the development. The original thermal envelopes used in the construction of the MCR dataset were compiled through examination of a variety of published distribution maps, museum records and collection reports spanning c. 150 years of collection data (Perry, 1986). The reliability of this data is variable, and there is a need to reassess and supplement them with more modern data. In addition, it is now possible, through the increase in commonly available computer power, to match collection dates against high resolution climate data and improve on the accuracy of the envelopes.

The expansion of the dataset to include phytophagous species would expand the usefulness of the method into other aspects of entomology, although because distribution may be limited by host plant their inclusion might cause theoretical problems in the reconstruction of Quaternary climates. The option to exclude these species from reconstructions would ensure backward compatibility with previous reconstructions.

There are a number of alternative systems for climate reconstruction from fossil insects evolving around the world (e.g. Huppert & Solow, 2004; Marra *et al.*, 2004; Bray *et al.*, 2006; Porch, 2006). Each method appears to have its own set of advantages and disadvantages, and it is unlikely that any single method will be universally applicable for all regions and timescales. The comparison of methods is essential, including a rigorous understanding of the quality of the primary data for each, in assessing the most appropriate methods.

^{ix} Much in the same way that ordination of fossil assemblages can only reveal grouping within the assemblage itself, and cannot provide an indication of the presence of these groupings in relation to habitats not represented by the fauna.

6 Case Studies – testing BugsCEP with real data

This chapter presents the application of BugsCEP to specific research problems. It compares the results and usefulness of the software with published modern and fossil work, and presents the analysis of this author's own investigations (Hemavan and Lockarp). Although BugsCEP is primarily aimed at palaeoecology, the testing of its features on modern data sets is essential in order to understand and validate the results produced by the BugStats system. The tools may also be of use to ecologists, especially when investigating faunal changes over time. The sites which are the topic of the primary discussions are shown in Figure 6.1, and readers should refer to the BugsCEP software for the location of any other sites mentioned.

The BugStats diagrams presented here have been tidied up, but the representation of the data is the same as the files exported by BugsCEP, and all results, unless otherwise specified, are reproducible through the program. There will be little discussion of confidence intervals or error margins associated with either the environmental reconstructions or the correlation coefficients calculated below. It is doubtful whether this would be meaningful for the reconstructions, given the semi-quantitative nature of the BugStats system (see Chapter 4). Confidence intervals for the correlation coefficients have been omitted for the sake of brevity, and due to the fact that these are only used as general indicators to support the patterns of sample similarities observed in the EcoFig diagrams.

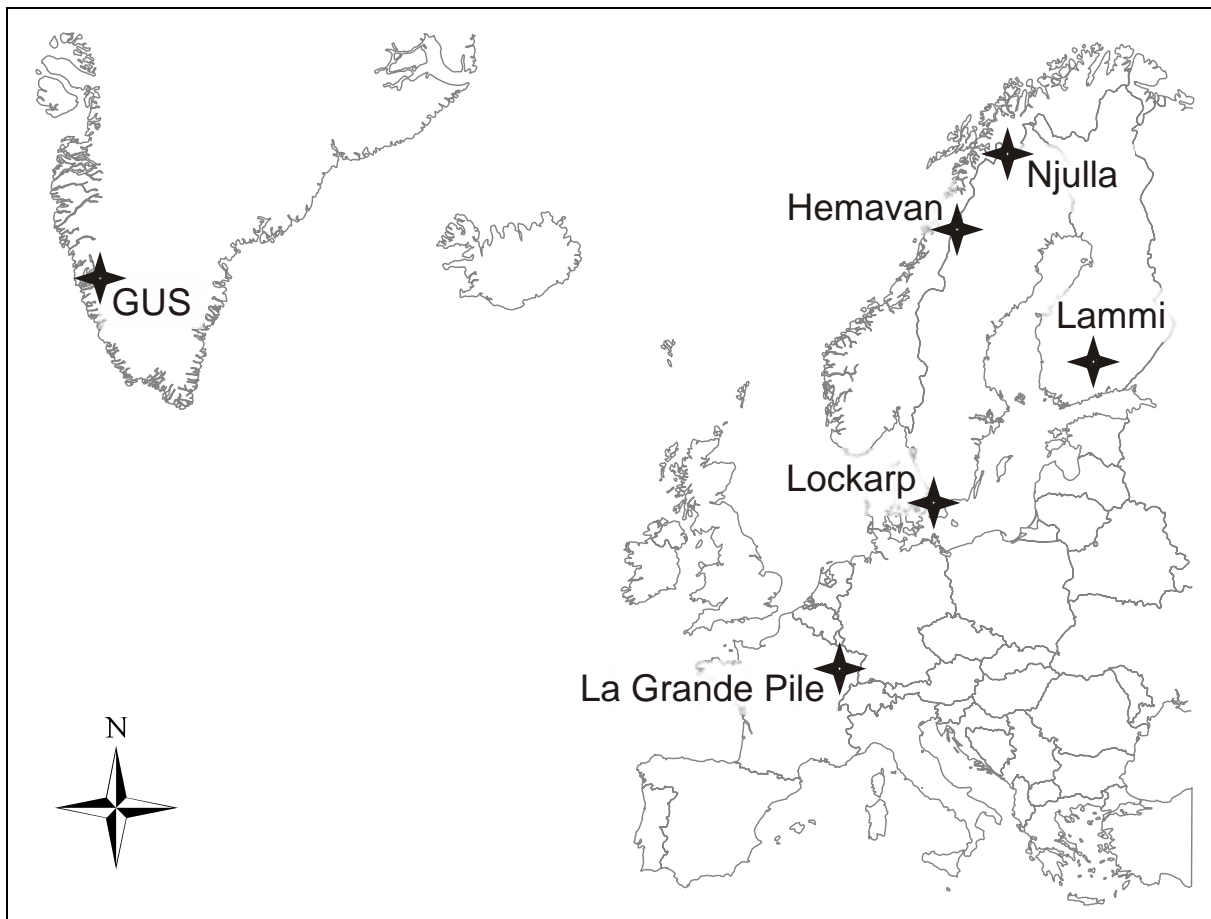


Figure 6.1. Map of case study sites discussed in this chapter. See BugsCEP for the location of other sites mentioned.

6.1 Modern Case Study: Forest-Farmland Pitfall Trap Transects in Lammi, Finland

Primary reference: Koivula *et al.* (2004).

6.1.1 Aims

Koivula *et al.* undertook a pitfall trapping project to investigate the carabid (ground beetle) communities on eight transects running from forest into farmland environments in the Lammi area of southern Finland (61°03'-61°08'N, 24°55'-25°05'E). They examined the frequency of species at different distances into each habitat, and discussed the implications of these findings with respect to a variety of habitat related traits for the species found. Their catch results, with species classified in three habitat groups, are presented in Figure 6.2, and are compared with the environmental reconstruction outputs produced by the BugsCEP component BugStats. The aim here is to assess the ability of the different BugStats calculation methods to reconstruct the modern habitats sampled in the original paper.

6.1.2 Introduction

Modern studies which use Coleoptera to examine environmental changes are, in principal, the same as those which reconstruct past environments. There are different taphonomic and identification problems, and the sampling methods are quite unlike, but the result is a set of species abundances from a number of sample points. Both studies require that the habitat preferences and environmental tolerances of the species identified are known. There is, however, a major difference in that modern studies have the definite potential to contribute to the understanding of the ecology of the species found, whereas fossil studies ultimately rely on other modern studies and can, perhaps, only contribute to modern ecology in terms of increasing the understanding the past biogeography of the species.

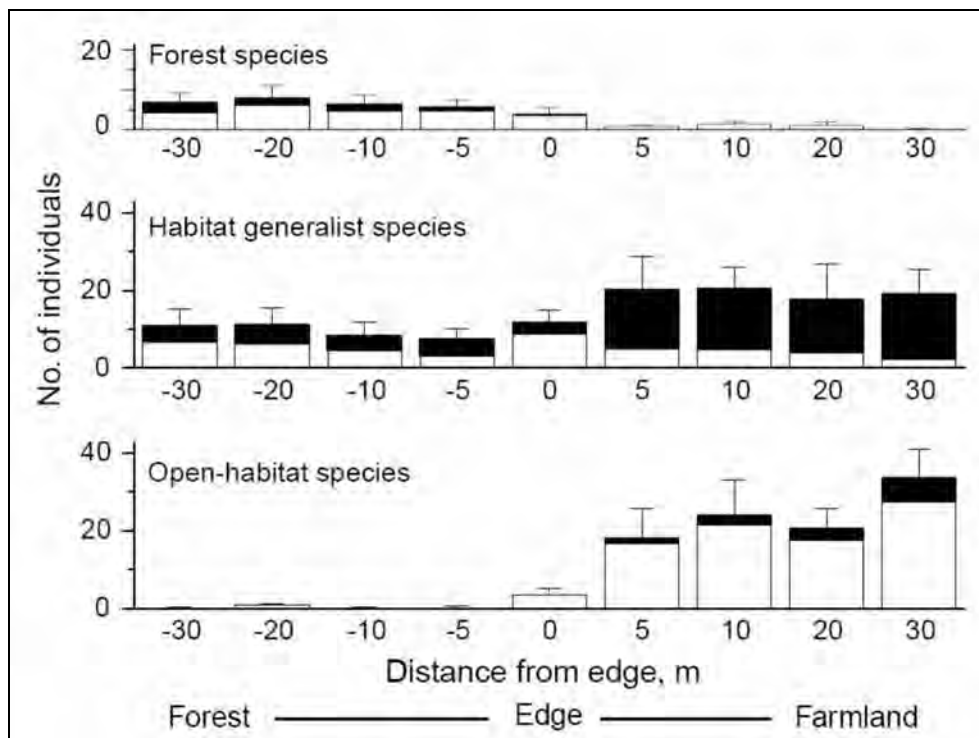


Figure 6.2. Mean adjusted catch (and SE) of carabids of three habitat groups along forest - farmland transects. Black/white shading represents different trapping episodes. Note that the Forest species y-axis has been rescaled in this thesis to match the other habitat scales. [Reproduced with permission from Koivula *et al.* (2004)].

Although designed primarily for assisting the study of past environments, the BugStats system can only be shown to be robust if it works for modern assemblages. It is, after all, built on the modern understanding of species habitat preferences. This case study tests that hypothesis for a small test set.

6.1.3 Methods

Koivula et al. (2004) methods:

Fifty nine carabid species were collected from a total of 288 pitfall traps along eight forest-farmland transects, as described in Table 6.1. Collection was undertaken in two periods between 24th May and 15th August 2001. Species were classified “...according to their coarse habitat association...” (*ibid.* p299) into Forest, Generalist, and Open habitat species. This is much the same method as used in the Bugs EcoCode classification system, although the information sources used vary somewhatⁱ, and BugsCEP includes a larger number of classes. Catch frequencies were adjusted to compensate for lost traps (mean no. of individual/trap x 20 days), and the grouped classifications presented as plots against the trap distances from the forest-farmland edge (Figure 6.2). The two most common species in each habitat class were: Forest: *Calathus micropterus* (Dft.) and *Pterost. Oblongopunctatus* (F.); Generalist: *Pterostichus melanarius* (Ill.) and *Pterostichus niger* (Schall.); Open: *Amara montivaga* Sturm and *Carabus cancellatus* Ill.; a full list can be found in the appendix of Koivula *et al.* (2004). Koivula *et al.* also undertook a number of statistical analyses (including ANOVA and DCA) in order to try to understand the response of different groups of carabid species to variations in the habitats sampled. These analyses will not be discussed here, but their conclusions will be referred to where relevant.

Table 6.1. Sampled landscape units summarized from Koivula *et al.* (2004)

Landscape unit	Description
Forest	<i>Oxalis-Myrtillus</i> -type mesic mixed-wood forests, dominated by Norway spruce (<i>Picea abies</i>), birches (<i>Betula</i> spp.) and aspen (<i>Populus tremula</i>). The age of the dominant trees was 80–120 years.
Farmland	Five grass (<i>Trifolium</i> and <i>Taraxacum</i>) dominated hay meadows; Two set-aside plots ((<i>Trifolium</i> and <i>Taraxacum</i> plus <i>Stellaria media</i>); One crop field (bare ground and wheat (<i>Triticum</i>) in respective sampling periods).

BugsCEP methods:

As the raw pitfall trap data were not published, three pseudo-samples were created to contain the grouped trap total abundance values. Koivula *et al.* provided total catch values for each species, and percentage catch values for three trap groups based on distance from the forest-farmland edge (Koivula *et al.* 2004: p307 - Appendix) as follows:

Forest: traps 10-30 m from edge

Edge: traps ± 5 m from the edge

Farmland: traps 10-30 m from edge

The adjusted catch percentage values were converted to approximate raw abundance values to enable abundance weighting in the calculations, and imported into BugsCEP as a countsheet attached to a new site. The BugStats EcoFig routine was run on this data with several calculation combinations. A standard BugsCEP full site report and species-EcoCode breakdown report are included in the digital appendix of this thesis for further reference. Detailed explanations of the calculations involved in the construction of Bugs EcoFigs can be found in Chapter 4, and will therefore only be summarized here:

ⁱ See the ‘Lammi’ digital appendix data on the CD accompanying this thesis for full details.

- Abundance weighted – the EcoCodes for each species are multiplied by the species sample abundance.
- Species presence only – the EcoCodes for each species are only counted once for each habitat class (also referred to as *no abundance*).
- Raw – count data are presented without standardization.
- %SumRep – count data for each habitat class are given as a percentage of the sum of the count data for all classes. These diagrams are more viable for inter-site comparisons.
- $\ln(n+1)$ – count data (+1) is natural logarithm transformed. This is done after abundance weighting and before applying %SumRep standardization, and can help compensate for skewed population distributions. (Note that Koivula *et al.* used the same transformation before performing ANOVA and DCA analyses on the data).

A species may represent more than one habitat class, and so abundance and NSpec (species richness) totals are not the same as the sum of environmental representation counts (SumRep) for each sample.

Eight different Bug EcoFigs can be created for data with species level identifications, six of which are presented here. $\ln(n+1)$ transformed ‘species presence only’ diagrams have been omitted as they are more informative of the degree of habitat specialization of the species involved, and of the code designation system itself, than of the environments represented by the species. The diagrams used collectively present different possibilities for the reconstruction of the environment from the list of beetles found in each trapping zone. Variations in the species compositions of each zone, when which are processed in the construction of each diagram, transform the habitat information in the data to signals of different relative strengths, which must be interpreted with care.

A BugsMCR thermal reconstruction was run on the data for the sake of experimentation, and as a further validation of the MCR dataset. The results are presented in Table 6.3.

6.1.4 Results and preliminary discussion

On first examination, none of the BugStats outputs (Figure 6.3) seem to illustrate the difference between the three habitat-association groups as clearly as the diagram presented in Koivula *et al.* (Figure 6.2). As the latter is from mean adjusted catches, one would expect the raw abundance weighted BugStats diagram to be most comparable (Figure 6.3b). This appears to be the case when comparing Koivula *et al.*'s Farmland with the Bugs EcoCode categories *Disturbed/arable* and *Sandy/dry disturbed/arable*ⁱⁱ, with the BugStats output practically mirroring the pattern created by the average of the transect traps. The drier nature of the farmland is clearly evident in the *Sandy/dry disturbed/arable* graph. *Heathland & moorland*, and *Meadowland* species are also more strongly represented in the Farmland traps. *Wood & trees*, however, is not clearly differentiated between the samples in this diagram (Figure 6.3b). However, the *Wetland/marshes* signal is stronger in the Forest sample, as would be expected, given the generally higher humidity and soil moisture usual in forest environments (Geiger *et al.*, 2003). A casual glance at the *Abundance* and *SumRep* graphs suggests that this could be due to the far greater abundance in the farmland traps, which is reflected in the sample sum of represented environments (SumRep) when using abundance weighting. In this case, the arable land tends to support both higher numbers of carabid individuals and species than woodlands, although this should be expected to vary with the nature of the substrate. It should also be noted that this is reflected in the number of carabid species respectively classed in BugsCEP as *Wood & trees* (82) and *Disturbed/arable* or *Sandy/dry disturbed/arable* (105). General ecological knowledge also tells us that ecotones (Edge) and disturbed environments have higher expected species richness than stable environments (Colinvaux, 1973). Although the Edge samples had higher species richness than the stable Forest environment, it was lower than the Farmland, suggesting that the effect of farmland disturbance on species richness is stronger than the ecotone (Edge) effect. This lends more support for the standardization of EcoCode graph values to enable inter-sample comparison.

ⁱⁱ BugStats classifications will be given *italicised* to avoid confusion.

By using the ‘%SumRep’ options in BugStats, graph values can be recalculated as percentages of the total sums of representations for each sample, thus compensating for differences in sample abundances (or species richness when abundance weighting is not used). When weighted abundance values are applied, the *Wood & trees* signal appears much more prominent (Figure 6.3a). A clear gradient can be seen in its drop through Edge to Farmland samples, and in addition, the soil moisture indicative classes *Meadowland* and *Wetland/marshes* give a stronger signal in the Forest traps. The two *arable* BugStats classes undergo little change except for the increase of the *Disturbed/arable* signal in the Edge sample with respect to the Farmland sample. It may be interesting to note that *Heathland/moorland* species are represented equally in the Forest and Farmland samples, and only slightly less in the Edge sample. This could either reflect a tendency in heathland Carabidae to be less habitat specific; be a reflection of the broad definition of the terms in the literature upon which the Bugs EcoCodes designations are based; the nature of the ground flora or substrate in the woodland; or be a product of the classification system.

Although the use of %SumRep standardization on the abundance weighted data increases the visibility of some expected habitat signals, it can be seen from the remaining diagrams that some information is missing. In particular, Figure 6.3a and Figure 6.3b do not show any indication of *Open wet habitats* or *Halotolerant* species (i.e. salt enriched environments), but the remaining four diagrams show a small signal. Before discussing the significance of the signals themselves, their variable visibility should be explained. Log transforming species abundancesⁱⁱⁱ is a well known tool for normalizing population distributions in both ecology (Jongman *et al.*, 1995; Krebs, 1989; Southwood, 1978) and palaeoecology (e.g. Hammer & Harper, 2006; Kenward, 1978). This reduces the positive tail common in beetle assemblages, and may help to give the population distribution a more symmetrical (normal) form that is a prerequisite for a variety of statistical methods (e.g. regression, ANOVA). In abundance weighted diagrams it also reduces the significance of high abundance species and raises that of the rare species. This is reflected in the BugStats output by a reduction in the relative difference between the highest and lowest values, effectively flattening the diagram components. By comparing Figure 6.3b and Figure 6.3d, we can see that although the aforementioned small magnitude signals are revealed, the visible difference amongst the higher magnitude signals is apparently reduced – making differentiation of the three habitat groups difficult from the diagram. In this case the transformation was not useful, and this may be the case for many sites; BugStats was designed to work with non-transformed abundance or species presence data.

The ‘Species presence only’ diagrams are, in this case, more useful than the log transformed ones. In order to create these, BugStats has ignored all abundance data and only looked at the presence(/absence) of species, effectively reducing the significance of super-abundant species and increasing that of rare species. This may help compensate for differential susceptibility of species to trapping methods, or differential taphonomic loss of fossils. In some respects, it has a similar effect to log transformation, but tends to preserve more of the habitat characteristics of sample values – especially when %SumRep standardized, as can be seen in Figure 6.3e. It is good practice to present both abundance weighted (cf. number of individuals) and non-abundance weighted (species presence/cf. number of taxa) results to give a fuller picture of the population structure (e.g. Ponel 1995).

ⁱⁱⁱ $\text{Ln}(\text{abundance}+1)$ is used rather than $\text{ln}(\text{abundance})$ due to the presence of zero values in most datasets.

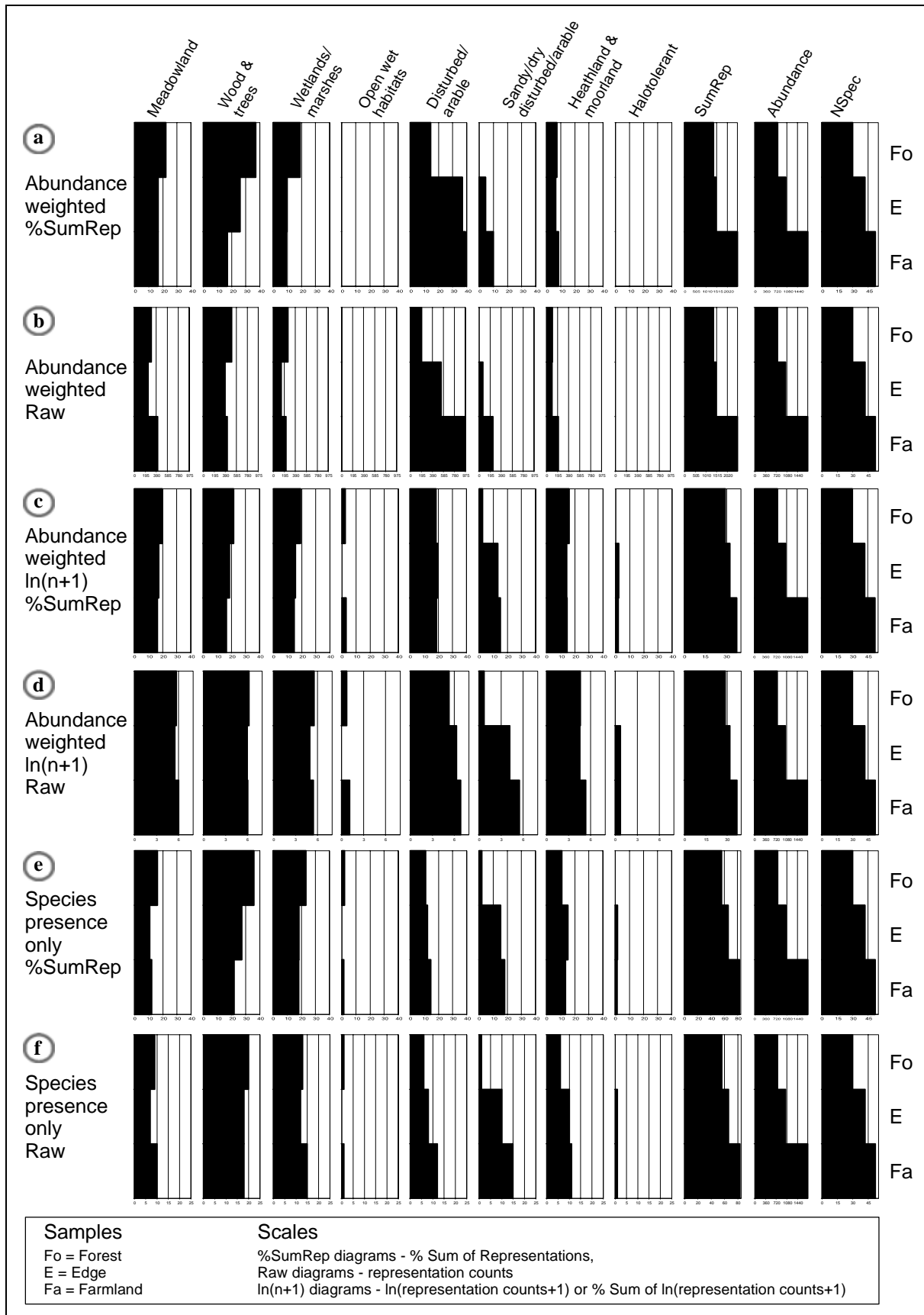


Figure 6.3. BugStats EcoFig output (edited) for Koivula *et al.* (2004) reconstructed grouped trap data. See text for discussion of variations. Note that the x-axes are different for each raw diagram. NSpec = Number of species.

Open wet habitats are to be expected in small forest openings, along stream banks or by pools and dew ponds in arable land. In this case, the signal is generated by a single occurrence of *Asaphidion flavipes* (L.) in the Forest, and two individuals of *Pterostichus vernalis* (Panz.) in the Farmland. *A. flavipes* is a eurytopic but hygrophilous species which prefers open environments (Koch, 1989; Lindroth, 1985), and is most probably an accidental presence in the Forest environment. *P. vernalis*, on the other hand, is a stenotopic, hygrophilous species common in moist areas of a range of environments (Koch, 1989; Lindroth, 1974; Luff, 1998), and so may indeed reflect a real, but small, microhabitat in the Farmland. *Halotolerant* species are most common in saline environments such as the shores of brackish water bodies, but may also be found in situations where high evaporation or erosion, be it natural or anthropogenic, increases the level of salts in the topsoil. They are therefore often present in arable landscapes, as ploughing mixes raw sediment with the upper organic soil horizons. *Halotolerant* species are totally absent from the Forest samples, and closer inspection of the fauna shows that one individual of the often coastal *Amara tibialis* (Payk.) (Koch, 1989; Luff, 1998) represents this habitat group in the Edge and Farmland samples. The *Halotolerant* habitat class should, however, be interpreted with care, as the reliability of the designation of a number of often coastal species is debatable and in need of revision.

Many authors (e.g. Kenward, 1978) often choose to exclude species with abundances lower than a certain threshold from analyses. This reduces the risk of casual occurrences affecting the results, but also risks filtering out small indicator and microhabitat signals that could contribute valuable information to an investigation. Such small scale indicators may or may not be important, and their usefulness will depend on the aims of the project^{iv}. A number of species are commonly only found in low frequencies (Colinvaux, 1973), and such filters potentially remove their input, and so should be applied carefully, with careful reflection on the sampling method.

Koivula *et al.* found that the Edge fauna was more similar to the Farmland fauna than the Forest Fauna, and the BugStats coefficient of similarity output supports this (Table 6.2).

Table 6.2. Modified Sørensen's coefficient of similarity (Southwood, 1978), showing slightly closer similarity between the Edge fauna and the Farmland fauna, than the Edge fauna and the Forest fauna^v.

	Forest	Edge	Farmland
Forest			
Edge	0.517165		
Farmland	0.300628	0.66289	

The MCR thermal reconstructions produced results comparable to the current central-southern Finnish climate (Table 6.3). A number points are of particular interest:

1. The reconstructions show a slightly wider TValue span in the Forest than in the other habitats, and indicate a slightly more continental climate in its TRange span. This has implications for the use of MCR with fossil material, in that the past environment must be reconstructed in terms of its habitats whenever calculating MCR, so that such factors can be considered. It can be suggested, from the results of this study, that reconstructed winter temperatures are more sensitive to differences in the habitat preferences of the species used for reconstruction, although more examples must be looked at before drawing further conclusions.

^{iv} It should be noted that the Bugs EcoCodes include five indicator classes, which have slightly different implications from the general habitat classes (see Chapter 4).

^v Significance testing of coefficient values is not yet implemented in BugStats, and these results should be treated as general approximations until tested.

2. The Edge habitat reconstruction values reflect its spatial position and species composition, i.e. in between the other two habitats, but more similar to the Farmland, which is supported by the correlation coefficient values (Table 6.2). The reconstruction of the winter climate also reflects this to a limited extent.
3. The reconstructed TMax span is narrower than the present day (1971-2000) span in central-southern Finland, whereas the reconstructed TMin span is wider than the present day span. The reconstructed temperatures are compatible with the recorded values, within the level of accuracy expected from MCR, but this clearly displays the lower resolution of winter temperature reconstructions.

Table 6.3. MCR thermal reconstruction for Koivula *et al.* (2004) reconstructed grouped trap data. All groups produced 100 % overlap regions. Present day climate for southern Finland given for comparison^{vi}. See Chapter 5 for explanations of the reconstruction method and variables.

Sample	TMaxLo	TMaxHi	TMinLo	TMinHi	TRangeLo	TRangeHi	NSPEC
Forest	15	18	-15	4	12	30	20
Edge	16	18	-14	1	16	30	26
Farmland	16	18	-14	-1	19	30	35
Present day climate (1971-2000), July and February respectively							
	13.7	20.9	-7.7	-2.2	Helsinki (60°10'N, 24°56'E)		
	10.4	21.3	-12.9	-5.0	Jyväskylä (62°24'N, 25°40'E)		

6.1.5 Further discussion

The selection of standardization and transformation options is evidently extremely important, and it is recommended that *all* output options are tested, and that the implications of each are understood thoroughly before interpreting the diagrams. A critical difference between the Bugs EcoCode classifications and the habitat groups used by Koivula *et al.* is that the former allow a species to be assigned to more than one class. Broadly eurytopic species (Generalists) therefore, may represent several environments, whereas stenotopic species will represent only one. This is why the sum of represented environments (SumRep) is always higher than the associated sample abundance or number of species (NSpec). Only one species in the dataset (*Loricera pilicornis* (F.)) was indicative of both *Woodland and trees* and *Disturbed/arable*, a designation which most likely represents the commonality of shaded habitats within forests and tall crops (Lemdahl, *pers. comm.*). Such microhabitat factors are difficult to cater for in general habitat classification systems, and the example serves to demonstrate the need for detailed entomological knowledge when interpreting results based on classification. A comparison of the respective groupings (summarized in Table 6.4, see digital appendix for full species comparisons) shows a reasonable agreement, with only three stenotopic species (with only one Bugs EcoCode classification) having possible conflicts between the systems.

BugStats includes indicator classes for *Deciduous* and *Coniferous* woodland – neither of which was indicated by the fauna. In other words, none of the Carabidae collected in this experiment are considered, as far as the BugStats system is concerned, as woodland indicator species at the European level. At the moment there are nine species of carabid classed as *Deciduous* indicators, and only two classed as *Coniferous*. It could be that the Bugs EcoCodes require revision in this area, although sources such as Lindroth (1945, 1985, 1986) should include a good understanding of North European coniferous forests, even if Koch's (1989-92) data are more Central European and deciduous. Alternatively, the explanation could be more ecological, in that the Carabidae are more tied to the

^{vi} TMax/TMin limit equivalent values. The transects (at 61°30'-61°80'N, 24°55'-25°50'E) most likely experienced a climate between the values from these stations. Data from Finnish Meteorological Institute (FMI) http://www.fmi.fi/weather/climate_6.html

environment of the woodland than the actual wood and trees themselves. That is to say they thrive in the shade and humidity provided by the forest cover, and in some cases prey on other species that are more specifically tied to the actual trees. An ideal study of the forest-arable interface should include as many Coleoptera taxa as possible to reduce the risk of generic level habitat preference traits biasing the results.

Table 6.4. Comparison of number of species per group/classification.
Note that BugStats allows species to fall into more than one class.

Number of Species per habitat group/class			
Koivula <i>et al.</i>		BugStats	
Forest (F)	13	Wood and trees	23
General (G)	12	Meadowland	11
Open (O)	34	Wetlands/marshes	18
Sum	59	Disturbed/arable	13
		Sandy/dry disturbed/arable	16
		Heathland & moorland	11
		Other	3
		Sum	95

6.1.6 Conclusions and implications for BugStats

One might expect the BugStats system to dilute the habitat signals of eurytopic species, as these species count values are spread among several classes. Such species give ‘false’ positive signals of the other habitats that they *could* survive in, in addition to the one that they currently occupy. This is of particular relevance when considering the differences between the interpretation of modern and fossil datasets. When dealing with fossil assemblages this reflects the reality of environmental reconstruction, in that we cannot know *exactly* which of the habitats are being reflected, and we can only use professional judgement to derive the most probable reconstruction. With modern datasets it is often possible to go out and survey the contemporary sampled environment, if it has not already been done in connection with the sampling. By complementing the fossil insect analyses with investigations based on other proxy sources (e.g. pollen, plant macro fossils etc.), one can often improve on the reconstruction by refining the probability of certain environments having existed.

It is a common misconception that post-modernism and empirical science are incompatible, the argument being something as follows: ‘As the past is not directly measurable, and its interpretation is a product of the subjectivity of the researcher, then a number of possible interpretations are equally valid. Hence we can not possibly know *anything* about the past and *all* interpretations are equally valid’. Readers will notice the considerable jump in logic from sentence one to two, which, it should be noted, is more often made by archaeologists and not the philosophers and historians of science from which it is derived. The ‘truth’ of the matter is that several possible interpretations are valid for any reconstruction, and a good publication will present this fact, either by describing a number of possibilities or presenting a measure of uncertainty or probability around the chosen interpretation. Suffice it to say that this is a discussion that has been ongoing within geography and archaeology for many decades (see e.g. James, 1972; Trigger, 1989), and in its essence for many centuries (see e.g. Simmons, 1993). Archaeology, ironically, appears to have a limited ability to allow for a plurality of ideas based on empirical evidence, but this may of course just be a phase that it is going through (Yoffee & Sherratt, 1993)^{vii}.

^{vii} It is perhaps a more serious matter when these misconceptions on our ability to interpret the past influence government policy on research and museum funding, as may be in danger of happening in Sweden (see e.g. Sundborg, 2006)

The slight difference in reconstructed temperatures from the Forest and Farmland habitats reinforces the importance of the effects of large scale vegetation on microclimates. This implies that habitat reconstructions should be studied along with MCR results, as different habitats may differentially bias regional reconstructions. However, the temporal and spatial resolution of palaeoenvironmental samples may reduce such effects, due to the fact that samples will generally have collected species from a wider variety of habitats and climates than samples from modern habitat specific studies. In addition, the evidence of a single study is insufficient evidence to draw further conclusions, and more modern dataset will have to be analysed from a wider range of habitats.

The reconstruction created by BugStats gives a reasonable reconstruction of the habitats described by Koivula *et al*, and possibly gives an indication of the surrounding species catchment area. This case study helps to reinforce the viability of BugStats for palaeoenvironmental reconstruction, by illustrating its ability to reconstruct reasonably accurately a known modern environment. As with any semi-quantitative method, the errors involved in the reconstruction are difficult, if not impossible to calculate. Some Quaternary scientists would perhaps claim that this reduces the power of the method when compared to other quantitative methods. I would disagree, and counter claim that many quantitative environmental reconstruction methods are in fact just as semi-quantitative as BugStats, and that the use of advanced statistical methods to provide standard errors does not elevate them automatically to a higher level of realism. The use of resampling methods on the other hand, which can be applied to semi-quantitative methods, to calculate resampled standard errors is a valuable technique for assessing the relative reliability of reconstructions with respect to the faunas that underpin them.

6.2 Fossil Case Study: 140 000 year Peat Sequence, La Grande Pile, France

Primary reference: Ponel, P. (1995).

6.2.1 Aims

Ponel (1995) studied the beetle fauna from approximately 13 m of multiple cores from a 19 m deep, 140 ka year long peat sequence from the Grande Pile peat bog in eastern France. The ombrotrophic bog is 25 ha in area, and lies at 325 m above sea level. In an impressive piece of work, Ponel identified 394 taxa, including 19 species no longer present in France, from 41 samples, and undertook an environmental and climatic reconstruction from this data. Ponel grouped species by habitat in much the same way that Bugs EcoCodes do, although the systems were developed independently^{viii}. This, and the presentation of his results as environmental indication diagrams displaying number of taxa and species (reproduced here with the permission of the author), makes the site an ideal case study for comparison with BugStats and BugsMCR outputs. In particular, the importance of standardization is apparent in some of the differences between Ponel's figures and those produced by BugStats.

For further descriptions of the site, samples, and other analyses from La Grande Pile readers should refer to the original paper. Similarly, there is no need to repeat detailed descriptions of the faunal changes here, apart from where they aid the discussion on the comparison of classification comparisons.

6.2.2 Methods

It was necessary to add a number of more southern taxa to BugsCEP to enable the entry of the Grande Pile list, and these new taxa have no climate data in the database, which could explain some of the differences between Ponel's MCR reconstruction and that produced by BugsMCR. The reasonable

^{viii} The classifications used in BugsCEP are influenced by his work, among others.

numbers of taxa and individuals present in many of the samples allowed for a wider variety of BugStats settings to be tested. Plots were produced both with limiting the calculations to only those individuals identified to species level as well as for all taxa. This allowed a greater degree of data interrogation, the limited plots theoretically providing more trustworthy reconstructions in that the more broadly categorized generic identifications (sp., spp., indet. in BugsCEP) are omitted. The implications of the differences between these plots are discussed below, along with both a general environmental interpretation focussing on the similarities and differences between Ponel's account and that produced by BugStats.

Ponel classified the Coleoptera into "...several categories according to their ecological requirements..." and, although BugsCEP does not have direct equivalents for all of his categories, the respective classes are shown in Table 6.5 for comparison. In addition, Ponel listed the number of taxa dependent on a number of specific plant taxa, which will not be compared here. Such analyses require only standard BugsCEP information retrieval (browsing and site reporting), and not the use of EcoCodes, which do not hold that level of detail.

Table 6.5. Comparison of Ponel's ecological requirement categories and Bugs EcoCode equivalents. Bugs EcoCodes that are not directly equivalent to Ponel classes are shown in *italics*.

Ponel classes (Ponel, 1995, p7)	Nearest Bugs EcoCode equivalent
Aquatic	Aquatics
- Standing-water	Indicators: Standing water
- Running-water	Indicators: Running water
Terrestrial	All except the above
Riparian species (that live on the shores of standing-water or on river banks)	<i>No direct equivalent, but possibly a combination of Wetlands/marshes and Open wet habitats</i>
Coprophagous (directly dependent on mammal faeces)	Indicators: Dung, and possibly Pasture/Dung and Dung/foul habitats.
Coprophilous (indirectly dependent on mammal faeces)	Dung/foul habitats, and possibly Pasture/Dung
Necrophagous species (that live on dead animals)	Carrion
Tree-dependent species, and deciduous/conifer dependant.	Wood and trees, Indicators: Deciduous and Indicators: Coniferous, respectively
Plant detritus feeders	<i>No direct equivalent, but possibly a combination of Dung/foul habitats and Mould beetles</i>

It is important to remember that while Ponel presents numbers of taxa and individuals, BugStats presents numbers of environmental representations, that is to say counts of the number of taxa that represent an environment. These values may be abundance weighted, or not, and optionally standardized, producing four different plots from the same data (excluding the log transformation option). It is also possible to limit the BugStats calculations to only taxa identified to species level, thus potentially increasing the reliability of the environmental indications displayed. All eight of these plot possibilities are presented below, but only selected components are described with reference to Ponel's work.

6.2.3 Results and general comparison

The Grande Pile fauna produced some kind of signal in all Bugs EcoCode categories, with the exception of *Stored grain pests* and *Ectoparasites*. The latter is hardly surprising due to the rarity of these finds in natural deposits, their concentrations, at least in present day Europe, being too few without natural herds or human concentration of animals. There is, however, very little known about

the extent of herbivore herds in Europe in the past^{ix}, and investigations of potential watering holes would probably be necessary in order to even tentatively identify them from the insect faunas. Finding these deposits is, unfortunately, potentially even more difficult. Pests of stored products are almost equally rare in natural deposits, with perhaps the exception of direct sampling of packrat middens (Elias, 1994) or the burrows and nests of other animals that collect seeds.

The presence of synanthropic species is interesting, and may be a reflection of a North European bias in the database's sources. It may also indicate that what we consider as synanthropic species today have previously been a consistent part of the natural background fauna. Although filtering out generic level identifications almost entirely removes these (Figure 6.4), it would be interesting to see what proportion of these species could be captured by pitfall trapping on the bog today.

Besides the groups discussed below, a number of categories produce almost continuous background signals, especially in the 'All taxa' diagrams (Figure 6.4). These include: *Meadowland*, *Heathland/moorland*; and to a lesser extent: *Open wet habitats*, *Disturbed/arable*, *Sandy/dry disturbed/arable*, *Halotolerant*. At least two of these, *Heathland/moorland* and *Open wet habitats* are explainable in terms of the local bog environment, and their persistence after the removal of generic level identifications reinforces this. *Meadowland* species may reflect a number of possible environments including naturally grazed grasslands, which would be supported by the *Pasture/dung* component, or other open landscapes with flowers and grasses. The other categories may, to an extent, be artefacts of generic level identifications representing a larger spectrum of habitat possibilities.

The effects of filtering the calculations to species level only vary between habitat group and sample. Some of the effects on groups will be the inevitable consequences of the removal of taxa from other groups, thus raising the percentage representation of the others. The sum of representations (SumRep) is calculated on the filtered counts, however, so this effect is internal to the particular diagram set (for a discussion of identification and taphonomic influences on environments represented, which will not be discussed in detail here, see Chapter 4). A number of general patterns are immediately obvious when comparing the 'All taxa' (Figure 6.4) and 'Species level identifications only' (Figure 6.5) diagrams:

1. Peaks and troughs are more pronounced in the species level only %SumRep than in the equivalent for all taxa, despite the reverse being true for the abundance and taxa sums (Figure 6.6). This may be an indication of a relationship between abundance and preservation, in that samples with higher abundances may also have better preservation, and *vice versa*.
2. A number of categories increase their % representation dramatically, with an increase by up to 65 % in some cases: *Aquatics*, *Indicators: Standing water*.
3. Some categories decrease their % representation significantly, becoming intermittent minor signals in the profile: *Open wet habitats*, *Meadowland*, *Pasture/dung*.
4. Some categories almost completely disappear: *Carrion*, *General Synanthropic*, *Halotolerant*.
5. A few categories remain essentially the same, as small, intermittent signals: *Indicators: Dung*, *Dry dead wood*.

By standardizing the data it is possible to obtain a picture of environmental change that is less dependent on the number of taxa or individuals in each sample. This is not without its interpretive implications, and readers should refer to the explanation of the standardization routine in Chapter 4 for more details.

^{ix} See the discussion of Frans Vera's work with respect to the insect evidence in Buckland [*et al.*] 2005.

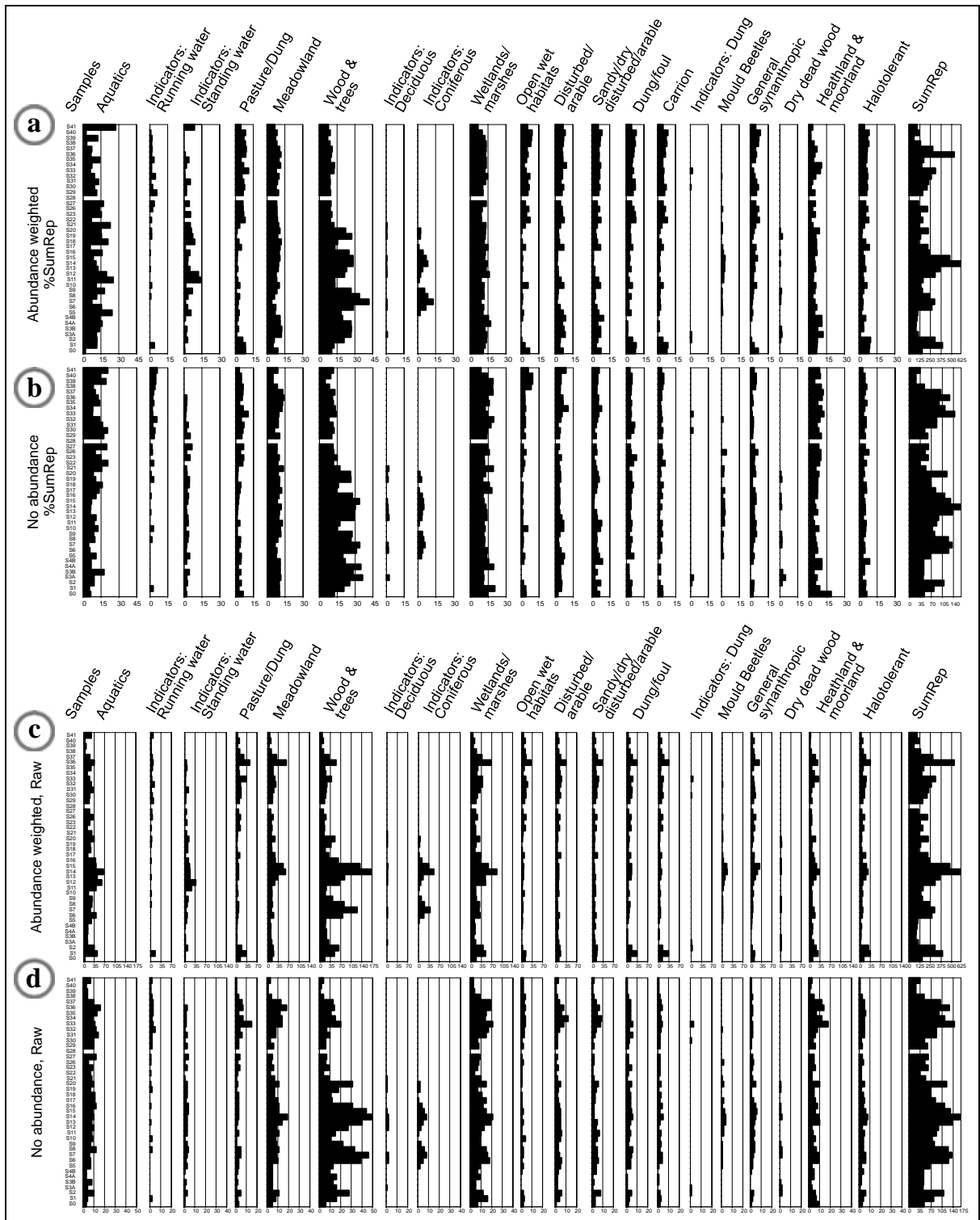


Figure 6.4. All taxa EcoFigs for La Grande Pile. Diagrams (a) and (b) show abundance weighted, and taxa only, standardized values respectively. Diagrams (c) and (d) show abundance weighted, and taxa only, raw values respectively. Number of taxa and abundance totals are shown in Figure 6.6.



Figure 6.5. Species identifications only EcoFig for La Grande Pile. Diagrams (a) and (b) show abundance weighted, and taxa only, standardized values respectively. Diagrams (c) and (d) show abundance weighted, and taxa only, raw values respectively. Number of taxa and abundance totals are shown in Figure 6.6.

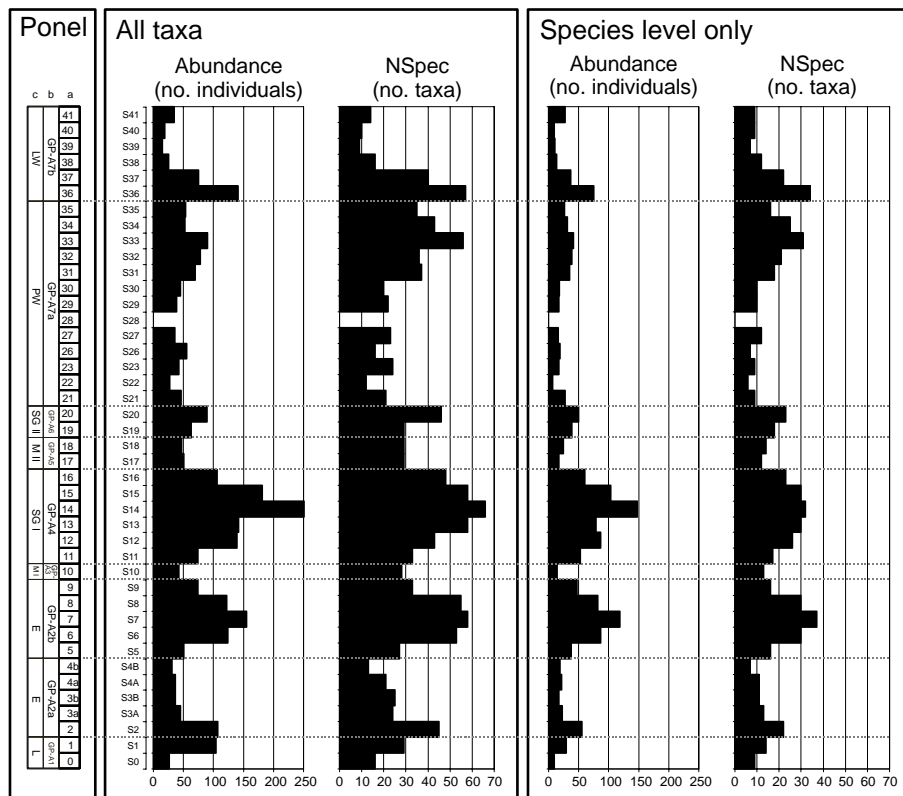


Figure 6.6. La Grande Pile, comparison of number of taxa and individuals, showing all taxa and species level only identifications for comparison. Ponel's (a) samples, (b) faunal units, and (c) the pollen chronozones of Beaulieu & Reille (1992) are shown. Note the different scales on abundance and NSpec graphs.

6.2.4 Discussion and detailed comparison of specific habitat groups

Ponel presents figures illustrating the variations in numbers of taxa and individuals that represent three particular habitat groups: tree-dependent species, aquatic species, and coprophagous species. A comparison of the equivalent classes (Table 6.5) in BugsCEP is presented here, in addition to a discussion of the problems with using raw abundance data for inter-sample comparisons, and the potential effect of standardization on interpretations. The BugStats plots which are most structurally similar to those of Ponel will be used as the basis for discussion, with other plots being presented where there are interesting differences, or to illustrate a point.

Note that the indicator classes in BugsCEP consist entirely of species level taxa, with the single exception of *Hydaticus* sp. This water beetle genus in particular is described as being totally dependent on standing water bodies, but it may be that other genera can be added to the indicator classes in later versions after discussions with the relevant experts.

It may be prudent to remind the reader of the long time scales involved with the interpretation of this data. With the insect samples covering approximately 110 000 years (see Ponel, 1995, for dates), the average time span covered by each sample is c. 2 700 years. Much can occur in such a space of time, including such difficult to entomologically define episodes as the Little Ice Age (Wagner, 1997), or the Younger Dryas, let alone the 8 200 cal. yrs BP event (Edwards *et al.*, 2006; see also Bennett, 2002, for a discussion of dating and correlation difficulties). As a result of this the faunas are likely to be mixed as the environment around the bog changes, and samples that overlap dramatically different phases, be they climatic or other environmental in nature, will provide a signal which indicates both of these environments. In the MCR results this will most likely result in a non-100 % overlap situation, which could theoretically be resolved by extracting a warm and cold faunal component from the data (see Chapter 5). Resolving the environmental signal is more difficult, as there are a number of classes involved, and doing so would necessitate even more circular reasoning than the MCR faunal

component concept. This said, the data still present a compounded environmental signal for each set of environments represented by the samples, which gives us valuable insights into the past 140 000 years of landscape change in the area around La Grande Pile. This in itself is one of the most complete records of long term environmental change that exists for Europe.

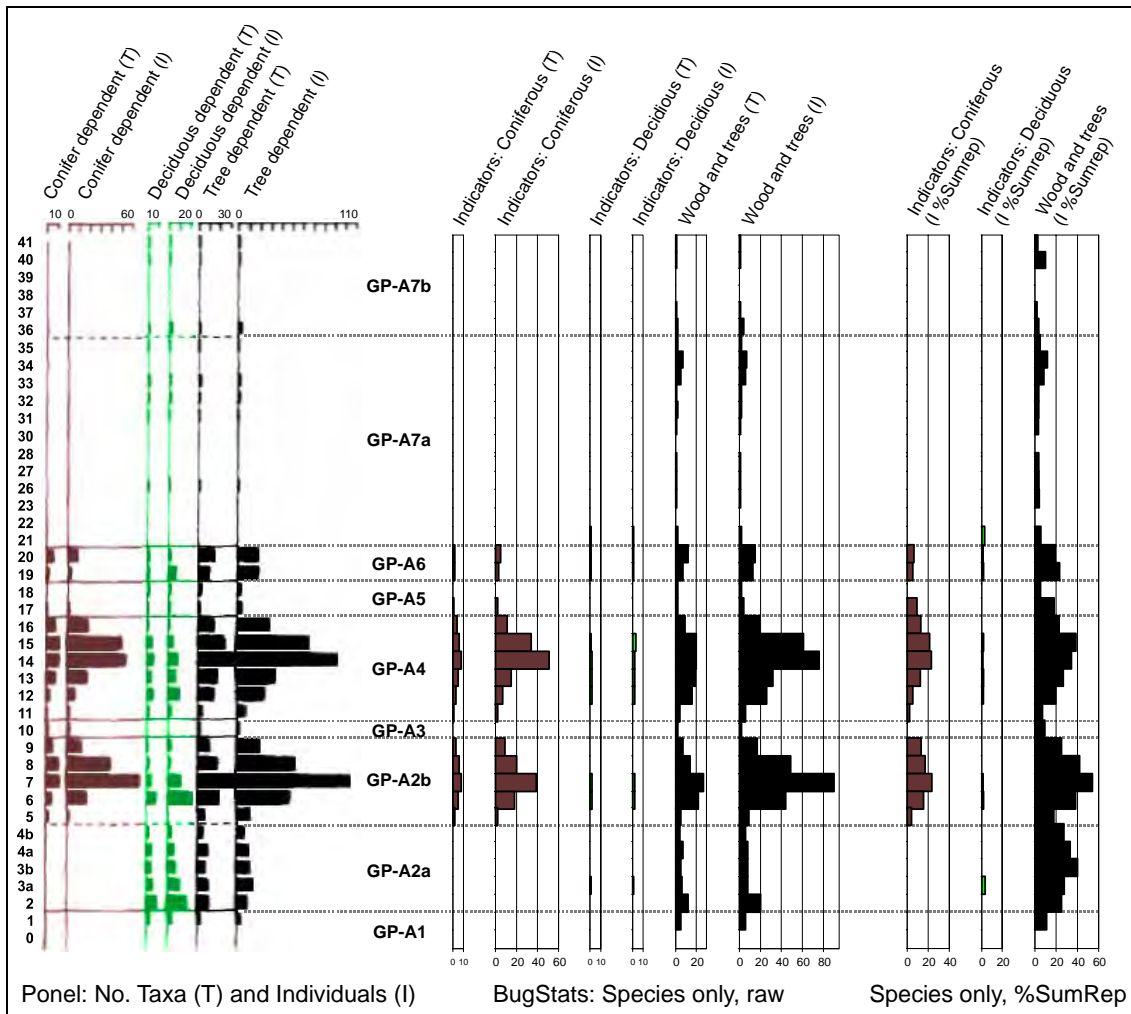


Figure 6.7. Comparison of woodland habitat numbers for Ponel's classes: Conifer dependant, Deciduous dependant, Tree-dependent; and the equivalent raw, species level identification only, representation counts in BugsCEP: *Indicators: Coniferous*, *Indicators: Deciduous*, *Wood and trees* respectively. T = No. of Taxa, I = No. of Individuals. Sample 29 is absent from the original figure and therefore also the BugStats charts.

The Tree-dependent and *Woodland and Trees* plots are almost identical between systems (Figure 6.7), including the standardized (%SumRep) plot to a certain extent. This shows a large degree of agreement in the classification of species by habitat. However, when it comes to comparing the indicator species with Ponel's deciduous and coniferous dependent groups there are a number of differences. BugStats does not show more than a single specific indication of deciduous woodland below sample 6, whereas Ponel shows deciduous dependent species in all but sample 0, with a peak in sample 2 almost equivalent to that in sample 6. There are in fact very few deciduous indicators in the sequence according to the BugsCEP classification, and there is obviously a significant difference in the classification systems in respect to this. This may be an aspect of the north European bias of BugsCEP, and the need for more data and a review of the classification system. It may also indicate that many of the taxa regarded by Ponel as being deciduous dependent are less so in northern Europe. The conifer indicators/dependent categories, although with slightly lower values in BugStats, show considerable similarity. There appear to be fewer *Wood and trees* species in BugsCEP than in Ponel's

tree dependent group, which is surprising, as the former class includes species less specifically tied to woodland, and more to the shaded environments provided by trees as well as those specifically dependent on wood or leaves. The inclusion of generic level identifications (Figure 6.4) gives this category a higher background signal, but does not change the structure of the peaks and troughs.

The %SumRep standardization of the abundance weighted species only data has some interesting implications for the interpretation of the woodland/tree categories (Figure 6.7). Whilst reducing the two most prominent peaks, it produces a peak between samples 1 and 4B comparable with that between samples 12 and 16. The implication of this is a much earlier woodland development in the sequence, at the start of Pone1’s faunal unit GP-A2a, rather than later in GP-2b.

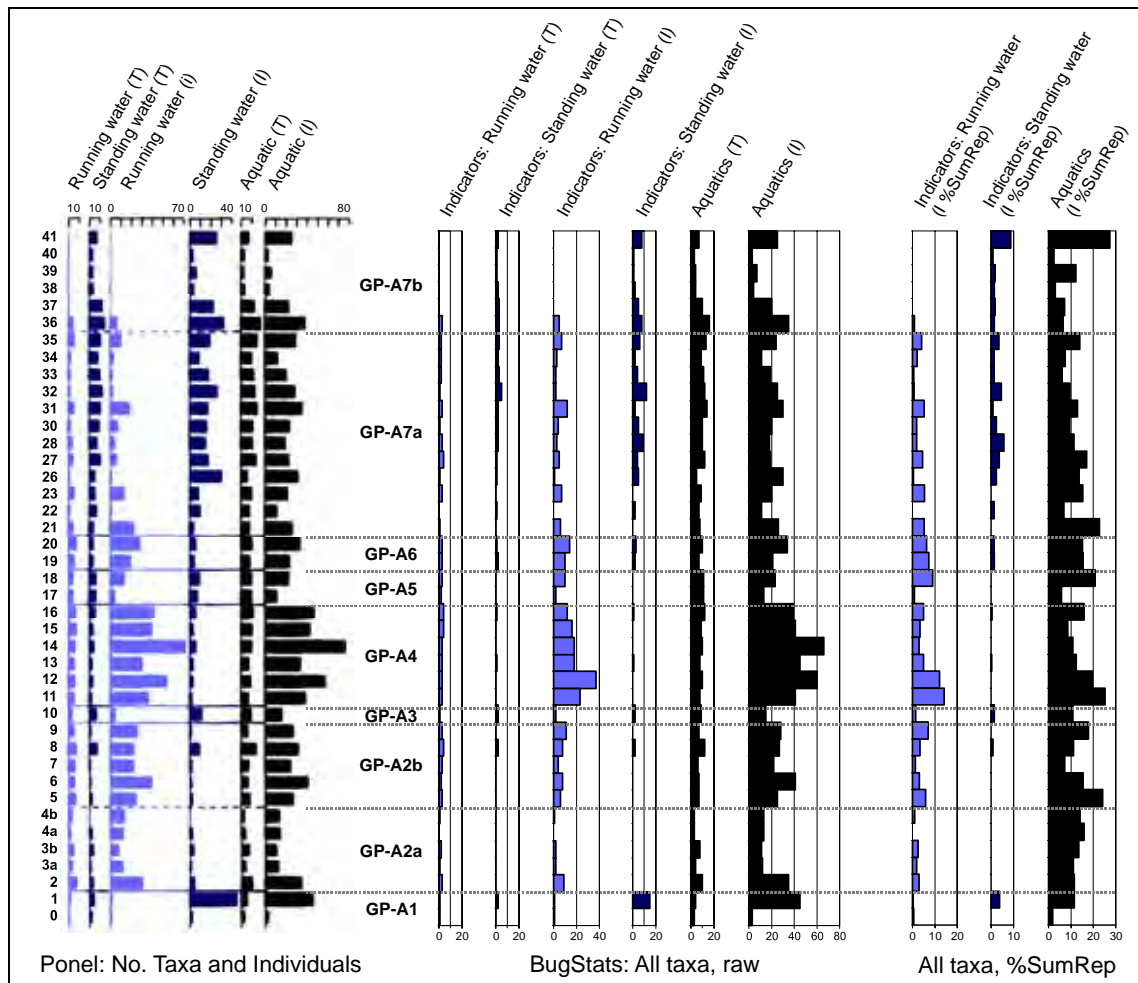


Figure 6.8. Comparison of aquatic habitat numbers for Pone1’s classes: Aquatics, Standing-water, Running-water; and the equivalent for all taxa in BugsCEP: *Aquatics*, *Indicators: Standing water*, *Indicators: Running water* respectively.

As with the woodland diagrams above, the general Aquatic groups match very well, both in number of taxa and individuals (Figure 6.8). This may reflect the fact that most water beetles are very easy to define as such, and their ecology is well known (e.g. Foster, 2000; Nilsson & Holmen, 1995). Although the trends are generally the same, the indicator type classes differ a little. This may primarily be due to a reluctance in BugsCEP to define species as either *Standing* or *Running water* dependent unless it is universally so for the BugsCEP dataset. In addition, generic level identifications have not been assigned to indicator groups, and it can be seen from the data that Pone1 uses the former in his reconstruction. As discussed earlier, the use higher level taxon as indicators may be valid for a limited region, but as Bugs EcoCodes are designed to be more regionally independent, then indicator classifications at the generic level are risky.

Standardization of the data makes some of the clusters visible in the raw data difficult to resolve. In particular, the lowest peak in aquatic individuals (samples 1-2) merges completely into the samples above it. As the total number of individuals in these samples is high (104 and 107 resp.), an explanation in terms of large percentage swings due to low numbers can be discounted, and one must look to the other habitats with which it is standardized for clarification. From Figure 6.4a (or indeed Figure 6.4c), we can see that although the number of aquatic individuals is high, there is an equally high signal in *Wood and trees* and *Wetlands/marshlands*, and almost as high in *Halotolerant, Heathland & moorland* and *Pasture/Dung*. This represents quite a varied landscape of woodland, bog, grassland and moorland with lakes or streams. This illustrates, of course, that whilst changes in abundances for one group of insects on its own is interesting, it must be put in a landscape perspective for a fuller reconstruction. There is a similar peak reduction effect at samples 35-37, although part of the peak, at 35, remains. This appears to be reflected in a slight decrease in the *Pasture/Dung* signal, possibly indicating a wetter period with generally fewer grasslands and grazing animals, as shown in Figure 6.9.

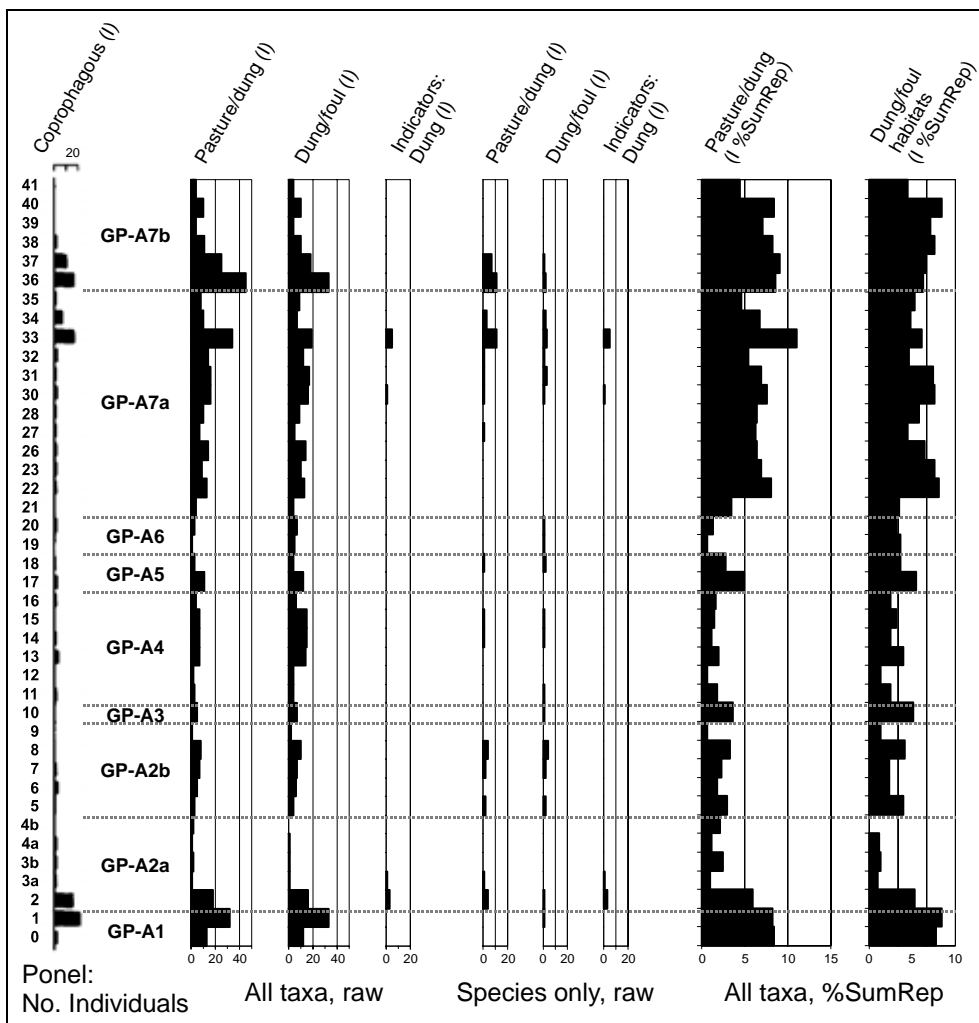


Figure 6.9. Comparison of dung habitat numbers for Poneil's coprophagous class and the equivalents in BugsCEP: *Indicators: Dung*, *Dung/foul habitats* and *Pasture/Dung*. Note the different scales.

BugsCEP includes three EcoCode categories that may give indication of the presence of dung: *Indicators: Dung*, *Dung/foul habitats* and *Pasture/dung* (Figure 6.9). These are designed to include a large part of the full range of coprophagous (feeding on dung) and coprophilous (found in dung) beetles. The specificity of habitat requirements of dung beetles varies, and so a dung beetle does not always definitely indicate the presence of dung. Some species are as likely to be found living in wet

mud, although many must breed in dung (Skidmore, 1991). Part of the need for these different classes is that archaeological sites can have had a number of ‘dung like’ deposits, such as middens, that attracted dung beetles without there necessarily being domestic animals present. The indicator class is logically the most restrained, and is the most definite indicator of dung, and thus animals in the surroundings, modern comparative work suggesting that even small numbers of some of these species are indicative of the presence of animals (Lemdahl, *pers. comm.*). Ponel’s number of coprophagous individuals diagram most closely resembles the species only, raw, *Pasture/dung* output from BugStats, with the exception of the lowest two-three samples. Closer examination of the fauna show that the latter is entirely due to the exclusion of *Aphodius* spp. from the BugStats diagram. It seems that the two classification systems do not completely mesh in this case, where although the genus *Aphodius* is largely a dung indicator, it does include several species (e.g. members of the subgenus *Aegolius*) which prefer rotting plant debris, and the uncertainty in the identification excludes it from the most comparable diagram. The ‘All taxa’ diagrams include counts for *Aphodius* spp., but are far broader than Ponel’s group, and thus indicate a greater variety of habitats.

The above may serve to illustrate the importance of examining the fauna in detail, even with the advantage of semi-automated habitat reconstruction software. The standardized diagrams (Figure 6.9), their scales expanded here for clarity, suggest a clear polarity in the sequence, with animals present in larger numbers at the bottom (0-2) and top (22-40/41), similar to that described by Ponel, but more extended. This is almost in direct antiphase with the woodland signals (Figure 6.7), but is closer to the open habitat signals (Figure 6.4). As noted by Ponel this corresponds closely with the colder phases of the sequence (Figure 6.10).

Although the precise details of the climate reconstruction system used by Ponel are not known, one can assume that the majority of the reference dataset is the same as that used in BugsMCR, which is part of the reason why the figures are so similar. There are a number of samples where the reconstruction of Ponel produces significantly more constrained temperature limits, especially for TMin: samples 19, 9, 4a, 3a and 1 are the most pronounced. In the bottom three of these the difference can be explained by less stenothermic envelopes defined in the BugsMCR dataset for low number of species (2, 2 and 1 resp.). These relatively warm tolerant species, *Poecilus lepidus* (Lesk.), *Agabus sturmii* (Gyll.), *Chaetarthria seminulum* (Hbst.) and *Coelostoma orbiculare* (F.) will have to be examined and their envelopes possibly revised. A number of other samples display narrower temperature ranges in the BugsMCR reconstruction, in particular, samples 37, 33 and 30. None of these samples have a 100 % overlap area, and it may be that Ponel has chosen to omit species in order to simulate one, rather than calculate temperature values from the nearest maximum area, as BugsMCR does when instructed.

The issue of overlap percent and sample representation is interesting, in that samples without 100 % overlap areas by definition must represent more than one, or a gradient along, temperature regimes. It may also be possible to extrapolate this further by suggesting that they may also represent more than environment and thus different sets of habitats. Although the function to extract warm and cold components from the fauna is not yet implemented in BugsMCR (see Chapter 5), the appropriate data and structure are in place, and it would be an interesting experiment in theoretical biogeography to reconstruct the environments represented by these components. Interestingly, sample 34, which has a comparable number of species to many of the non-100 % overlap samples, did produce a 100 % overlap reconstruction. This suggests an unusually stable temperature regime, or alternatively a much shorter time-span captured by the sample. There also appears to be no or very little relationship between reconstructed temperature and number of species^x, which is unusual, as more temperate climates usually support more species than arctic ones (e.g. Rosenzweig, 1995).

^x At most significant is with ln(TMin upper limit) at $R^2=0.2$.

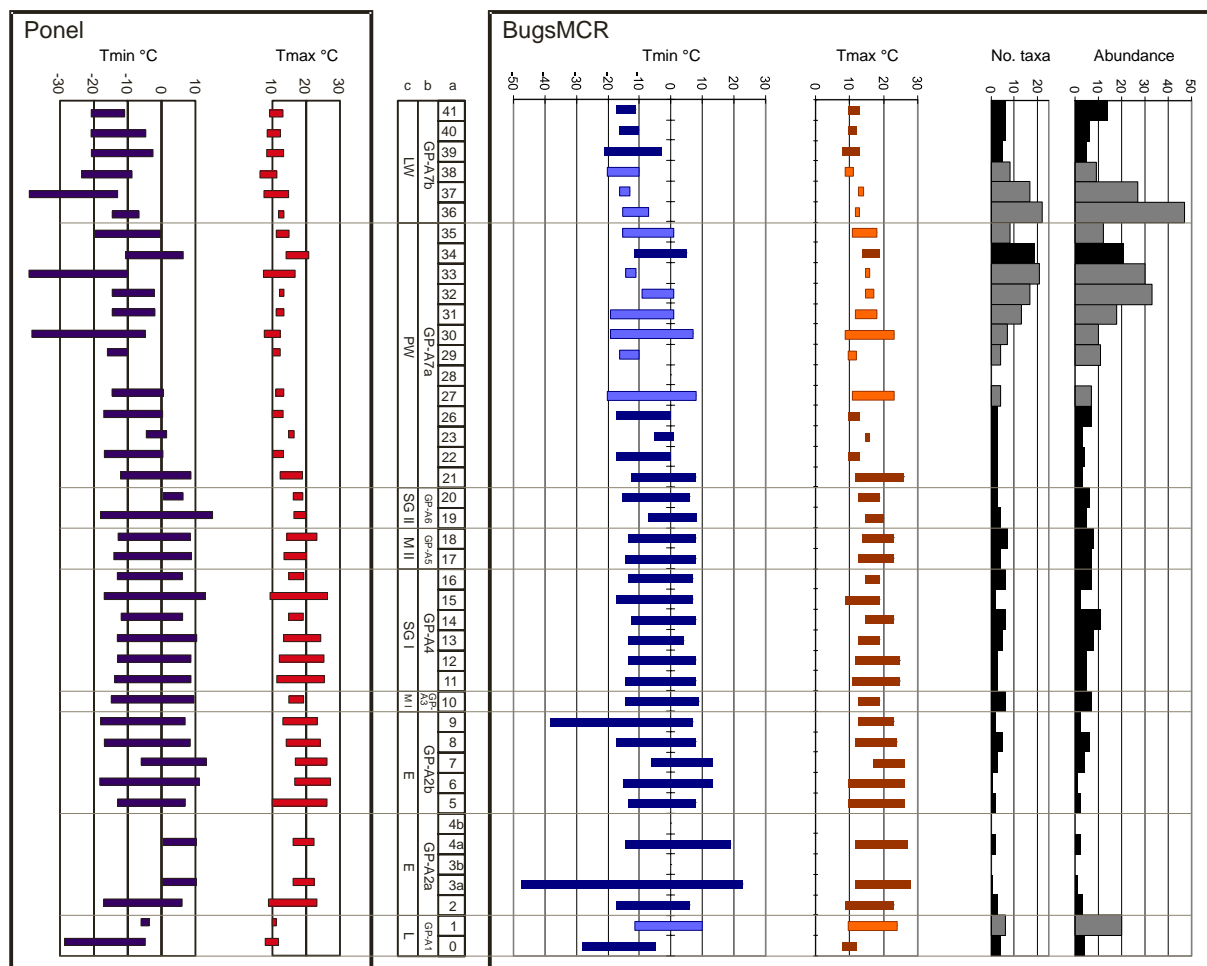


Figure 6.10. A comparison of the BugsMCR temperature reconstruction for La Grande Pile with that provided by Ponel. Numbers of taxa and individuals used in the BugsMCR calculations are provided on the right, and samples where the area of maximum overlap contains <100 % of taxa are indicated by lighter shading.

6.2.5 How are Ponel’s faunal and chronological groups affected by the use of standardized results and sample correlations?

Division of stratigraphic sequences into faunal units is a difficult task, and although a number of methods such as cluster analysis and ordination can be employed (Jongman *et al.*, 1995), the fine tuning of the division between units is as much an art as a science. Faunal units can be derived by statistically assessing the similarity between sample faunas, and then applying any of a variety of clustering algorithms to group samples by similarity and stratigraphic position (Kovach, 1995). A semi-quantitative approach, more common in palaeoentomology, is to identify significant changes in faunal signals by examining plots such as those used above, and with reference to, if available, MCR and other proxy data.

The BugStats correlation matrix output, comparing each and every sample fauna from La Grande Pile, is shown in Table 6.6. This could be used as the foundation for cluster analysis proper, but is used here to provide support for the statistical grouping of possible faunal units in the sequence. As can be seen there are a number of patterns in the similarity of neighbouring samples (represented by the values of the diagonal edge of Table 6.6), similar in many respects to those described by Ponel (1995). Coefficient values were used to group more like samples together, using an arbitrarily defined drop in similarity to define borders. Sequences of dissimilar samples were also grouped together, for the sake of simplicity, but the lack of similarity between adjacent samples suggests a high degree of variation in the faunas, and thus the environments represented by them (Figure 6.11). Ponel’s brief faunal unit descriptions (Ponel, 1995: p26) are quoted below for comparability. Refer to Figure 6.11 for the labels

used in the comparisons below, and previous figures for details of the habitat signals and temperature values.

Table 6.6. Correlation matrix showing modified Sørensen’s coefficient (Southwood, 1978) between La Grande Pile samples. Darker shading represents greater similarity between sample pairs, and coefficient values are on a scale from 0 = no similarity, to 1 = total similarity.

	S41	S40	S39	S38	S37	S36	S35	S34	S33	S32	S31	S30	S29	S27	S26	S23	S22	S21	S20	S19	S18	S17	S16	S15	S14	S13	S12	S11	S10	S9	S8	S7	S6	S5	S4b	S4a	S3b	S3a	S2	S1					
S41																																													
S40	0.41																																												
S39	0.40	0.41																																											
S38	0.23	0.55	0.40																																										
S37	0.31	0.38	0.27	0.38																																									
S36	0.20	0.20	0.13	0.20	0.52																																								
S35	0.31	0.27	0.20	0.30	0.40	0.36																																							
S34	0.05	0.03	0.03	0.13	0.23	0.27	0.22																																						
S33	0.10	0.18	0.11	0.23	0.34	0.36	0.26	0.35																																					
S32	0.23	0.23	0.11	0.29	0.35	0.40	0.39	0.24	0.33																																				
S31	0.10	0.22	0.12	0.25	0.28	0.29	0.52	0.24	0.25	0.42																																			
S30	0.23	0.38	0.13	0.34	0.40	0.32	0.48	0.16	0.28	0.42	0.38																																		
S29	0.35	0.38	0.15	0.34	0.32	0.28	0.41	0.20	0.29	0.46	0.29	0.55																																	
S27	0.31	0.25	0.20	0.30	0.34	0.28	0.47	0.22	0.24	0.35	0.32	0.49	0.43																																
S26	0.44	0.38	0.23	0.28	0.46	0.33	0.31	0.07	0.25	0.35	0.26	0.44	0.36	0.40																															
S23	0.18	0.26	0.21	0.24	0.27	0.21	0.42	0.17	0.24	0.23	0.34	0.39	0.32	0.38	0.31																														
S22	0.29	0.61	0.29	0.42	0.43	0.29	0.35	0.05	0.26	0.25	0.25	0.53	0.42	0.38	0.51	0.38																													
S21	0.15	0.15	0.16	0.14	0.13	0.12	0.20	0.06	0.09	0.16	0.12	0.20	0.19	0.22	0.16	0.39	0.22																												
S20	0.03	0.04	0.04	0.05	0.09	0.11	0.15	0.08	0.07	0.11	0.18	0.09	0.09	0.13	0.13	0.21	0.09	0.34																											
S19	0.02	0.00	0.00	0.02	0.06	0.08	0.10	0.05	0.01	0.06	0.12	0.02	0.04	0.08	0.07	0.19	0.00	0.33	0.53																										
S18	0.07	0.03	0.03	0.08	0.15	0.14	0.20	0.16	0.10	0.14	0.21	0.11	0.14	0.22	0.08	0.27	0.08	0.34	0.34	0.31																									
S17	0.14	0.26	0.15	0.32	0.32	0.21	0.34	0.17	0.20	0.29	0.33	0.27	0.33	0.30	0.26	0.30	0.28	0.25	0.21	0.16	0.35																								
S16	0.04	0.03	0.03	0.08	0.08	0.12	0.19	0.13	0.09	0.15	0.22	0.12	0.14	0.18	0.10	0.26	0.06	0.26	0.49	0.34	0.25	0.19																							
S15	0.05	0.05	0.04	0.05	0.07	0.10	0.12	0.06	0.07	0.11	0.17	0.09	0.12	0.10	0.08	0.18	0.07	0.25	0.43	0.32	0.19	0.16	0.56																						
S14	0.04	0.04	0.03	0.04	0.07	0.08	0.10	0.05	0.06	0.10	0.14	0.09	0.10	0.08	0.07	0.12	0.06	0.20	0.35	0.24	0.15	0.14	0.43	0.56																					
S13	0.05	0.05	0.05	0.07	0.11	0.13	0.12	0.09	0.10	0.12	0.16	0.06	0.10	0.12	0.10	0.17	0.07	0.27	0.50	0.39	0.30	0.23	0.54	0.53	0.48																				
S12	0.02	0.03	0.03	0.04	0.07	0.08	0.09	0.05	0.04	0.06	0.13	0.05	0.06	0.08	0.06	0.12	0.05	0.26	0.48	0.41	0.27	0.17	0.46	0.44	0.44	0.52																			
S11	0.02	0.02	0.02	0.08	0.07	0.09	0.09	0.09	0.06	0.08	0.18	0.05	0.07	0.15	0.06	0.17	0.04	0.32	0.47	0.41	0.40	0.18	0.38	0.29	0.26	0.45	0.51																		
S10	0.13	0.20	0.14	0.15	0.21	0.17	0.23	0.11	0.27	0.13	0.13	0.21	0.22	0.23	0.25	0.29	0.32	0.16	0.08	0.08	0.16	0.26	0.14	0.09	0.08	0.16	0.07	0.09																	
S9	0.02	0.02	0.02	0.02	0.05	0.07	0.14	0.05	0.04	0.08	0.14	0.07	0.09	0.09	0.08	0.21	0.06	0.47	0.52	0.54	0.35	0.19	0.49	0.43	0.35	0.47	0.45	0.43	0.10																
S8	0.04	0.03	0.03	0.03	0.05	0.11	0.11	0.08	0.08	0.12	0.18	0.12	0.11	0.13	0.07	0.22	0.04	0.26	0.31	0.28	0.17	0.16	0.32	0.36	0.34	0.30	0.31	0.26	0.11	0.41															
S7	0.03	0.02	0.02	0.03	0.06	0.08	0.10	0.06	0.06	0.09	0.08	0.08	0.08	0.07	0.14	0.04	0.21	0.33	0.27	0.14	0.15	0.34	0.37	0.33	0.35	0.24	0.21	0.07	0.35	0.51															
S6	0.03	0.01	0.01	0.04	0.05	0.06	0.08	0.05	0.03	0.05	0.08	0.04	0.05	0.05	0.04	0.13	0.04	0.24	0.38	0.31	0.25	0.16	0.33	0.33	0.32	0.40	0.33	0.31	0.04	0.33	0.28	0.45													
S5	0.02	0.03	0.03	0.03	0.05	0.06	0.10	0.10	0.03	0.03	0.13	0.08	0.09	0.07	0.02	0.17	0.05	0.35	0.37	0.33	0.24	0.16	0.29	0.25	0.23	0.31	0.29	0.43	0.06	0.38	0.32	0.30	0.34												
S4b	0.00	0.00	0.00	0.04	0.02	0.03	0.05	0.02	0.02	0.02	0.03	0.06	0.03	0.00	0.14	0.03	0.23	0.22	0.23	0.13	0.05	0.19	0.14	0.13	0.20	0.15	0.25	0.05	0.25	0.13	0.22	0.31	0.46												
S4a	0.00	0.00	0.00	0.03	0.05	0.04	0.09	0.04	0.03	0.05	0.06	0.10	0.08	0.08	0.04	0.20	0.06	0.31	0.29	0.24	0.21	0.14	0.20	0.16	0.16	0.22	0.20	0.22	0.05	0.27	0.14	0.22	0.30	0.36	0.50										
S3b	0.00	0.00	0.00	0.00	0.04	0.06	0.11	0.04	0.03	0.00	0.07	0.07	0.05	0.05	0.02	0.15	0.06	0.17	0.22	0.20	0.24	0.09	0.10	0.11	0.10	0.18	0.16	0.13	0.08	0.23	0.14	0.15	0.26	0.27	0.41	0.51									
S3a	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.04	0.02	0.00	0.00	0.16	0.03	0.26	0.24	0.24	0.15	0.04	0.12	0.14	0.13	0.17	0.17	0.15	0.02	0.22	0.16	0.21	0.31	0.33	0.42	0.56	0.56									
S2	0.01	0.02	0.02	0.05	0.10	0.09	0.10	0.11	0.14	0.08	0.12	0.07	0.08	0.08	0.06	0.17	0.04	0.22	0.22	0.24	0.22	0.16	0.24	0.24	0.21	0.33	0.23	0.28	0.09	0.25	0.15	0.24	0.31	0.29	0.29	0.26	0.22	0.22							
S1	0.17	0.20	0.07	0.17	0.27	0.28	0.25	0.11	0.28	0.25	0.17	0.27	0.31	0.20	0.42	0.18	0.29	0.08	0.08	0.01	0.08	0.21	0.08	0.06	0.06	0.12	0.04	0.06	0.23	0.04	0.06	0.05	0.06	0.05	0.03	0.07	0.07	0.04	0.20						
S0	0.20	0.58	0.29	0.47	0.34	0.22	0.28	0.10	0.28	0.25	0.29	0.42	0.43	0.29	0.37	0.32	0.53	0.14	0.07	0.02	0.05	0.31	0.06	0.06	0.04	0.08	0.02	0.04	0.18	0.02	0.04	0.06	0.04	0.03	0.04	0.03	0.03	0.00	0.09	0.25					

GP-A

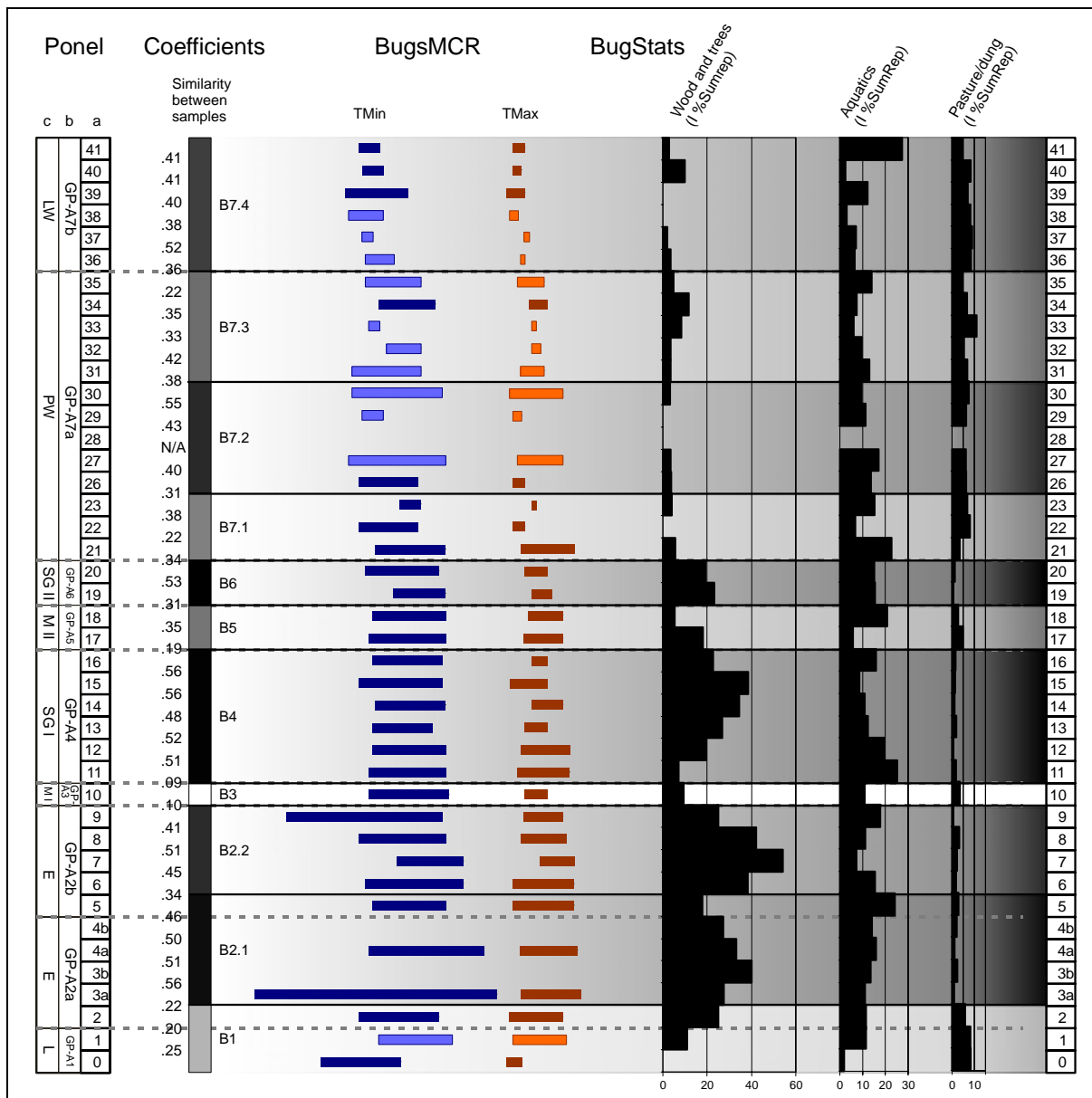


Figure 6.11. Comparison of Poneil's faunal units with those derived from correlation coefficients for neighbouring samples, and possible implications of BugStats standardized outputs on the original faunal units. Coefficient based sample groups (B1-B7.4) are shaded by the average of the coefficient values for samples in the group. 0 = no similarity, 1 = total similarity.

GP-A3:

“...a small unit showing a significant decrease in tree-dependent Coleoptera and the dominance of standing-water Coleoptera over running-water Coleoptera.”

This is apparent in the drop and narrowing of TMax, a reduction in *Aquatic* habitats, and a slight increase in *Pasture/dung*. The *Wood and trees* signal is ambiguous in that the graph depression continues into the sample above it. In addition, an examination of the other habitats available (Figure 6.4 and Figure 6.2) suggests a general spread of more *Open wet habitats*, *Disturbed/arable* and *Heathland & moorland* type environments, during a slightly colder period, as Poneil also notes.

GP-A4:

“...this unit is very similar to GP-A2b, but with sporadic occurrence of isolated specimens of the relatively cold-adapted species *Potamonectes assimilis*.”

BugStats reveals a lower woodland peak, closer in form and magnitude to GP-A2a than GP-A2b, but with a *Dung/foul habitats* more similar to GP-A2b. There is an interesting long term trend in the reduction of *Aquatics*, primarily generated by running water species (Figure 6.8). The correlation matrix (Table 6.6) shows a slightly higher statistical similarity with GP-A2b than GP-A2a.

GP-A5:

“...fairly similar to GP-A3, with a pronounced decrease of tree-dependent Coleoptera, a slight rise of standing-water beetles and a rare occurrence of cold-adapted taxa.”

A relative doubling of the *Pasture/dung* signal reinforces the definition of this unit, as does the MCR, although this shows a slight rise in possible temperatures despite the presence of cold-adapted species. BugsCEP does not produce any standing water signal for the unit. The standardized *Aquatic* and *Wood and tree* signals are ambiguous, mirroring each other in samples 17 and 18, with 17 having a relatively high *Wood and trees* signal, and 18 having an equally high *Aquatic* signal. The fauna of sample 17 appears to be statistically most similar to samples higher up in the sequence (18, 29, 31, 35, 37, 38) along with the basal sample. Sample 18 on the other hand, is more similar to the two samples directly above (19, 20) and a few samples below (9, 11, 13) in GP-A2b and GP-A4. It is most similar to sample 11, with which it shares very similar MCR results.

GP-A6:

“This unit has similar beetle assemblages to that recorded in units GP-A4 and GP-A2b. Tree-dependent Coleoptera reappear but are less abundant in GP-A6 than in GP-A4 and GP-A2b. There is a predominance of running water Coleoptera. Cold-adapted Coleoptera are rare.”

This unit is clearly visible in all three habitat groups discussed here (Figure 6.11), with woodland signals being elevated well above the adjacent samples. *Pasture/dung* and *Aquatic* signals are lower than the immediately adjacent samples, but the latter is still relatively high at c. 15 % SumRep. The *Dung/foul* signal (Figure 6.9) does not differentiate this sample from its neighbours, suggesting that whilst the grassland type landscape may have diminished, there were still muddy or relatively dung rich habitats nearby. The correlation matrix certainly illustrates that the samples are most similar, as Ponel states, to GP-A4 and to a lesser extent GP-A2b, although sample 9 is most similar to both 19 and 20. It is interesting to observe that sample 10, the probable open, cold phase unit GP-A3, bears almost no similarity at all to GP-A6. The *Aquatics* are dominated by running water species, but BugStats does show the highest proportion of standing water species since samples 10, although it is not much of a signal. The running water signal itself does not differentiate the unit from its immediate neighbours.

GP-A7:

“This large unit is made up of 18 samples. The beetle assemblage shows a great change compared with the lower samples. The tree-dependent taxa disappear almost totally, cold adapted and standing-water species increase in numbers and there is a corresponding decline in the numbers of running-water beetles. This unit may be divided into two subunits GP-A7a and GP-A7b, the latter is defined by an increase in cold-adapted species and an almost total loss of any running-water element.”

From the start of GP-A7a the environmental signals become more variable, and, as described above include a constantly higher *Pasture/dung* signal, which is more general than Ponel’s coprophagous group. There is general agreement between the systems in terms of woodland and water signals, and the GP-A7a/A7b split is compatible with both BugStat’s running water indicators and the MCR values. Although there is a reasonable amount of internal variation in A7a, it is tempting to suggest three subdivisions, as illustrated by B7.1, B7.2 and B7.3 (Figure 6.11):

B7.1: Mostly distinguishable by its middle sample, 22, which has no species level woodland signal at all, and a dramatically reduced aquatic fauna. It also produces the highest correlation value for the entire sequence, of 0.61 with sample 40. Climatically the unit appears little different from B7.2, although the smaller proportion of standing water species may indicate a different precipitation regime.

B7.2: Similar temperature regime to B7.1, but more constant representation of environments, and an increase in indications of standing water.

B7.3: A slightly warmer period, with greater numbers of taxa and individuals, and a slight increase in woodland habitats.

6.2.6 Conclusions

The intuitive classification of Ponel, based on entomological knowledge, compares closely with the Bugs EcoCodes classification developed independently from a variety of sources. BugStats output is remarkably similar to Ponel's diagrams, considering they are two independently developed systems. Ponel however, uses the number of species and individuals in his plot (closest to the 'raw' outputs in BugStats^{xi}), and the ability to standardize the BugStats output, in terms of percentages of environmental representations of each class by sample sums, allows for some differences in interpretation. As BugStats is intended to be a trans-European system, it inevitably generalizes more than any system built upon single site or regional investigations will do. There is also a significant central-north European bias in BugsCEP as a whole, and southern species may be under-represented and even under defined in terms of their habitat requirements.

Perhaps the most important lesson from this comparison is that whilst raw abundance data are useful for examining apparent changes in numbers representing particular habitats over time, it can be misleading when examining the relative changes over a number of habitat types. Changes in sample diversity may be masked by changes in numbers of taxa and individuals between samples, and the latter may be more a reflection of differential sample size, deposition rate and preservation. Standardization allows us to compensate for these variations without diluting the important habitat signals too much. If used in conjunction with carefully selected indicator species, this can be a powerful tool for interpreting complex faunal changes.

The assignment of species to habitat groups may require further discussion and this is an ongoing process, and such an endeavour is an interesting prospect for future work.

6.3 Fossil Case Study: Early Holocene Environmental and Climate Change at Hemavan, Northern Sweden

Own data contribution: fossil insect data

Other contributors: pollen – Engelmark (1996)
plant macrofossils – Tegby (2004)



Figure 6.12. Panorama photograph of the Hemavan bog taken from the sampling location.

6.3.1 Aims

The lake eroded face of a bog near Hemavan, northern Sweden, was sampled for pollen analyses in the early 1970s by Engelmark (1996). This site was revisited in 2003 by the author of this thesis, and new

^{xi} But not identical, as described in Chapter 4, a BugsCEP taxon may be included in more than one habitat, as may its abundance value.

samples taken close to the original site for the analysis of fossil insects and plant macrofossils. The results are interpreted below with the help of the analytical tools built into BugsCEP.

6.3.2 Site, samples and methods

The Hemavan site lies at 65°50'N, 15°00'E, 450 m a.s.l. in the northern Swedish district of Storuman (Figure 6.13). The sampling location was on the exposed face of a slightly raised bog close to the Ume river, where it is slowly being eroded by wave action from the adjacent lake (Figure 6.12). The rate of erosion appears to have slowed in recent decades, as the location of Engelmark's (1996) sample column was easily identified, despite having been dug in the early 1970s. This, along with the dry surface and obvious ill health of the bog, may be evidence of a falling water table.

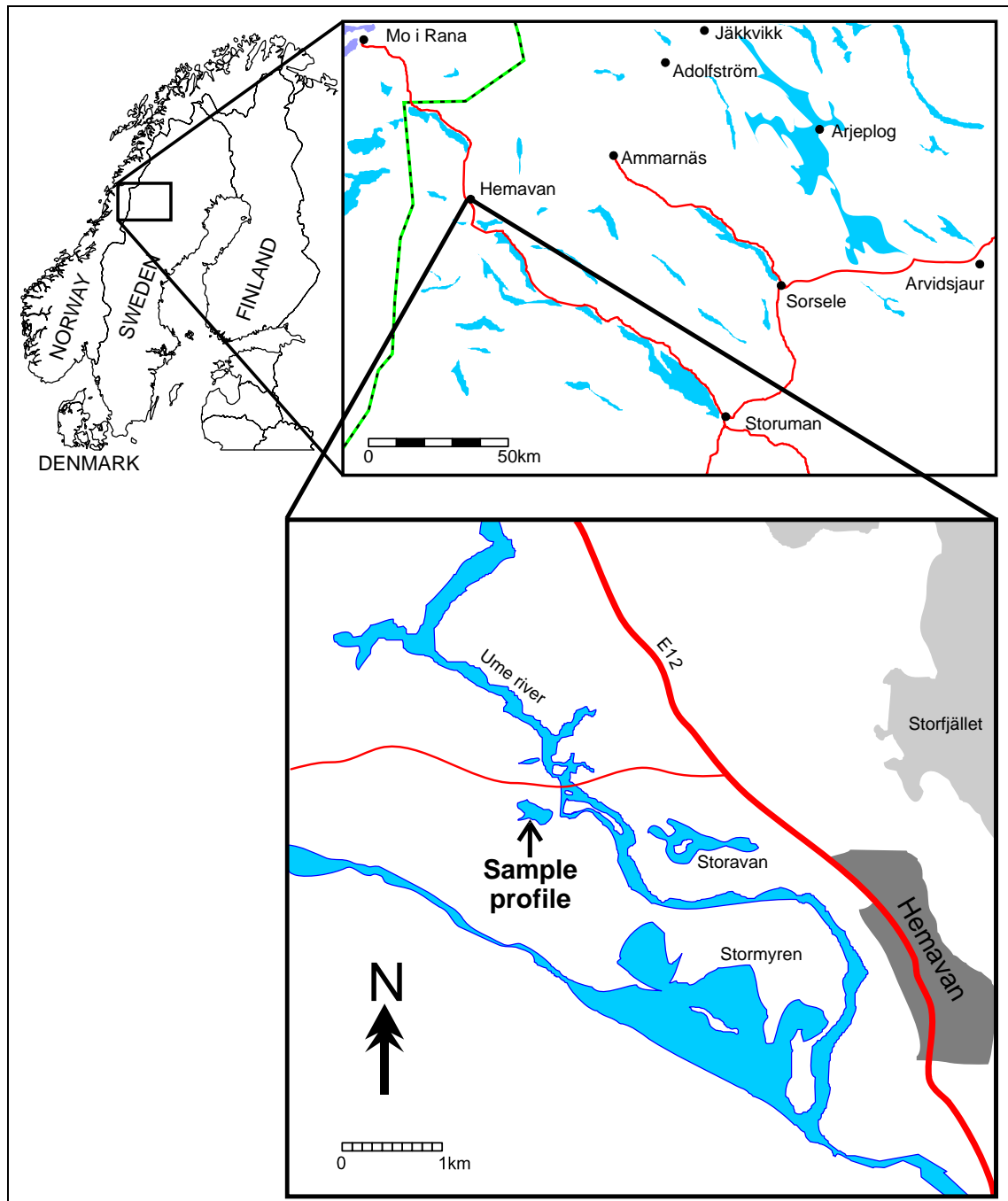


Figure 6.13. Map showing the Hemavan site and location of the sample profile.

Although there are no known archaeological sites in the immediate area of the site, there is considerable evidence of Stone Age activity around the large lake Överuman, some 7 km upstream (Holm, 1991). Previous work has shown that a bog or lake has to be very close to a hunter-gatherer occupation site for any record of the latter's impact on the landscape to be recorded (Hicks, 1993). It was not expected that the effects of the sites surveyed around Överuman would be seen in the samples, but any immediately local site of contemporary age might be visible in terms of faunal changes. It is, however, extremely difficult to differentiate between small scale human impact on the landscape (e.g. dead wood removal, burning) and natural environmental change.



Figure 6.14. Photograph of the sampling location, with remains of Engelmark (1996) section in the foreground on the left. Paul Buckland is documenting the vegetation (and simuliids) in the background.

It was decided to sample from the same profile as Engelmark (1996) (Figure 6.14), both to increase the comparability of results and to reduce damage to the bog. The section was cleaned, and a total of 30 samples of approximately 5 litres each were extracted, in 5 cm slices, covering a depth of 160 cm (Figure 6.15). The top 50 cm of peat were avoided due to obvious desiccation, which suggested that preservation would be poor. Extraction of the lowest 30 cm of sediment proved difficult due to ground water and arm length, and so the sample depths are less reliable for these samples (132 to 160 cm). Samples 130-135cm and 132-135+cm are from the base of the peat and the top of the clay respectively. Sampling in this way, rather than an even 5 cm block, should prevent the mixing of faunas from different sedimentological and environmental regimes. The sample column was approximately correlated with that of Engelmark (1996) by use of his raw data and distinguishable horizons (a dark humified band, the peat-clay interface) within the sediment (Figure 6.15), and the approximate depth from the peat surface.

Ten samples were chosen for analysis, to represent a reasonable chronological depth whilst focusing on the clay-peat transition. Unfortunately, time constraints prevented the analysis of further samples,

but they are in cold storage for future work. Sub-samples were taken for future reference (e.g. geochemistry), and the sediments described according to Troels-Smith's (Aaby & Berglund, 1986) classification for unconsolidated sediments (Table 6.7). Samples were disaggregated in warm water and washed through a 300 μm sieve. A few of the less humified samples swelled enormously (up to 200 % in volume) and were treated as two samples sieved at 300 μm and 2 mm respectively (the smaller catching the drainage of the larger). The insect faunas were examined separately from these and accordingly labelled as 'big' and 'small'. All samples were paraffin floated three times, washed in detergent, stored in ethanol and then examined under a binocular microscope. Species were identified by the use of reference literature and the modern reference collection at Manchester Natural history Museum (UK), by the author, with the assistance of Paul Buckland.

Abundance data was fed into BugsCEP as the site "Hemavan", and EcoFig and MCR plots produced. Plant macro fossils were also identified from number of samples by Ida Tegby (2004), and the results are discussed in relation to the insect data below.



Figure 6.15. Photograph showing the sampling pit at two stages of excavation. The scale is approximate and for illustration only. A dark, more humified band can be seen at c. 100 cm, and the peat-clay transition is visible at about 135 cm.

Table 6.7. Hemavan – processed sample depths and descriptions. Sample depths are in cm.

Bottom depth	Top depth	Processed size (L)	Size >300 (ml)	Flot size (ml)	Troels-Smith	Description
55	60	5	5500	75	Tb ¹ 4Th ¹ +Tl ¹ +	Moderately-well stratified moss peat, some twigs and plant bits, very poorly humified, some rhizomes/roots (vertical).
60	65	6.5	5250	300	Tl ¹ 1Th ¹ 2Dh1	Moderately stratified, fibrous peat, many twigs, a few roots, some large twigs. Flot was efloated and sorted as fine and coarse parts.
65	70					
70	75					
75	80					
80	85					
85	90	7	7000	50	Th ² 1Dh2Dg1	Moderately stratified, fine fibrous peat, some-many rhizomes, very few other roots (4th float made).
90	95					
95	100					
100	105					
105	110					
110	115	7	7500	100	Th ¹ 2Dh2	Well stratified, poorly humified peat, some roots.
115	120	5.5	7500	100	Th ¹ 3Dh1	Well stratified, poorly humified peat, very few twigs.
120	125	1+5	5000	60	Tl ¹ +Th ² 1Dh2Dg1	Poor-moderately humified peat, a few twigs ~fibrous (1 litre subsampled for unsuccessful NOH ₃ disaggregation test; 4th float made).
125	130	5	3500	100	Tb ² 1Th ¹ 2Tl ¹ 1	Poor-moderately humified peat, lenses of more humified peat. Some twigs and roots.
130	135	4.5	2000	50	Tb ³ 3Th ² 1+lenses	Moderately humified peat with lenses of less humified peat, and silt lenses. Uneven base to peat.
132	135+	4	700	50	95 %: As1Ag3Th ² + 5 %: lenses Tb ³ 3Th ² 1	Grey-brown clayey silt, many roots. Uneven top of clay.
135+	140	4	780	25	As1Ag3Dg ² +Dh ² +	Grey-brown clayey silt, some roots, fine organic fibres.
145	160	5	800	35	As1Ag3Dg ² +Dh ² +	Grey-brown clayey silt, some roots/reeds, fine organic material.

6.3.3 Results

Species abundances were generally poor, despite the large sample volumes (Table 6.8). The maximum sample abundance was 56 MNI (Minimum Number of Individuals), in the basal peat sample 130-135cm, and the minimum of two MNI occurred in the sample 135+-140cm. The most species rich sample was 60-65BIG (large float fraction), and this sample would most certainly have proved to have been exceptionally rich had there been time to identify the remaining fine float fraction.

Only standardized EcoFigs were produced (Figure 6.16), as the low abundance values in several samples would lead to un-proportional visual representation of the habitats of the species found.

Engelmark's (1996) pollen data are reproduced here in percentage form, along with selected parts of Tegby's (2004) plant macrofossil data for comparison. Although the numbers were small, enough MCR beetle species were present to allow for a rudimentary climate reconstruction (Figure 6.17) with some interesting results which are discussed below.

Table 6.8. Beetle species list from Hemavan (continued on next page).

Taxon	55-60Pt2	60-65BIG	110-115	115-120	120-125	125-130	130-135	132-135+	135+-140	145-160
<i>Loricera pilicornis</i> (F.)	1									
<i>Bembidion</i> sp.										1
<i>Patrobus septentrionis</i> (Dej.)						1				
<i>Patrobus assimilis</i> Chaud.				1			1			
<i>Patrobus</i> sp.		1								
<i>Pterostichus diligens</i> (Strm.)	1	1					3			
<i>Pterostichus strenuus/diligens</i> (Panz.)/(Strm.)				1						
<i>Pterostichus nigrita/rhaeticus</i> (Payk.)/Heer							1			
<i>Agonum piceum</i> (L.)		1								
<i>Agonum gracile</i> Strm.	1	1								
<i>Agonum fuliginosum</i> (Panz.)		1	1			1				
<i>Hydroporus memnonius</i> Nic.					1	2	4			
<i>Hydroporus</i> sp.		2								
<i>Hydraena britteni/riparia</i> Joy/Kug.								1		
<i>Cercyon</i> sp.							1			
Leiodidae indet.		1								
<i>Acrulia inflata</i> (Gyll.)		1								
<i>Olophrum consimile</i> (Gyll.)		5			1	3	3			1
<i>Olophrum rotundicolle</i> (Sahl.)	1						2			
<i>Olophrum</i> sp.		1								
<i>Eucnecosum brachypterum</i> grp.	5	2								1
<i>Acidota crenata</i> (F.)	1	1					1			
<i>Acidota cruentata</i> (Mann.)	1									
<i>Acidota quadrata</i> (Zett.)										1
<i>Anthophagus alpinus</i> (Payk.)		1								
<i>Anthophagus omalinus</i> Zett.	1									
Omalinae indet.		2					2			
<i>Bledius</i> sp.									1	
<i>Stenus</i> spp.	4	3			2	2	1			
<i>Euaesthetus bipunctatus</i> (Ljungh)		1	1	1						
<i>Lathrobium</i> spp.		6	1	2	3	6	8	1		
<i>Ochtheophilum fracticorne</i> (Payk.)		2	1		1	1				
<i>Philonthus</i> sp.	2	1					3			
<i>Platydracus pubescens</i> (Deg.)	1									
<i>Staphylinus erythropterus</i> L.						1				
<i>Quedius</i> spp.		4			2	1	2			
<i>Mycetoporus splendidus</i> (Grav.)		1								
<i>Tachyporus</i> sp.							1			
<i>Tachinus signatus</i> Grav.										1
<i>Tachinus</i> sp.					1					
Aleocharinae indet.	2	3					2			
<i>Bryaxis bulbifer</i> (Reich.)				1						
<i>Bryaxis</i> sp.	2	1	4	4	2	3				

Taxon	55-60Pt2	60-65BIG	110-115	115-120	120-125	125-130	130-135	132-135+	135+-140	145-160
<i>Pselaphaulax dresdensis</i> Hbst.							1			
<i>Rhagonycha testacea/femorialis</i> (L.)/(Brul.)							1			
<i>Rhagonycha</i> sp.							1			
<i>Hypnoidus rivularius</i> (Gyll.)										1
Scirtidae indet.	1			1	2					
Dryops sp.										1
<i>Cytilus sericeus</i> (Forst.)					2	6	4			
<i>Phratora</i> sp.						1				1
<i>Crepidodera fulvicornis</i> (F.)		1								
<i>Chaetocnema</i> sp.	1				1					
Curculionidae indet.							1			
<i>Otiorhynchus nodosus</i> (Müll.)							1			
<i>Bagous</i> sp.			1	2						
<i>Dorytomus taeniatus</i> (F.)								1		
<i>Ellescus bipunctatus</i> (L.)									1	
<i>Limnobaris dolorosa</i> (Goez.)			1	1	1	13	11			
Ceutorhynchinae indet.						1	1			
<i>Rhynchaenus testaceus</i> (Müll.)	1	1								

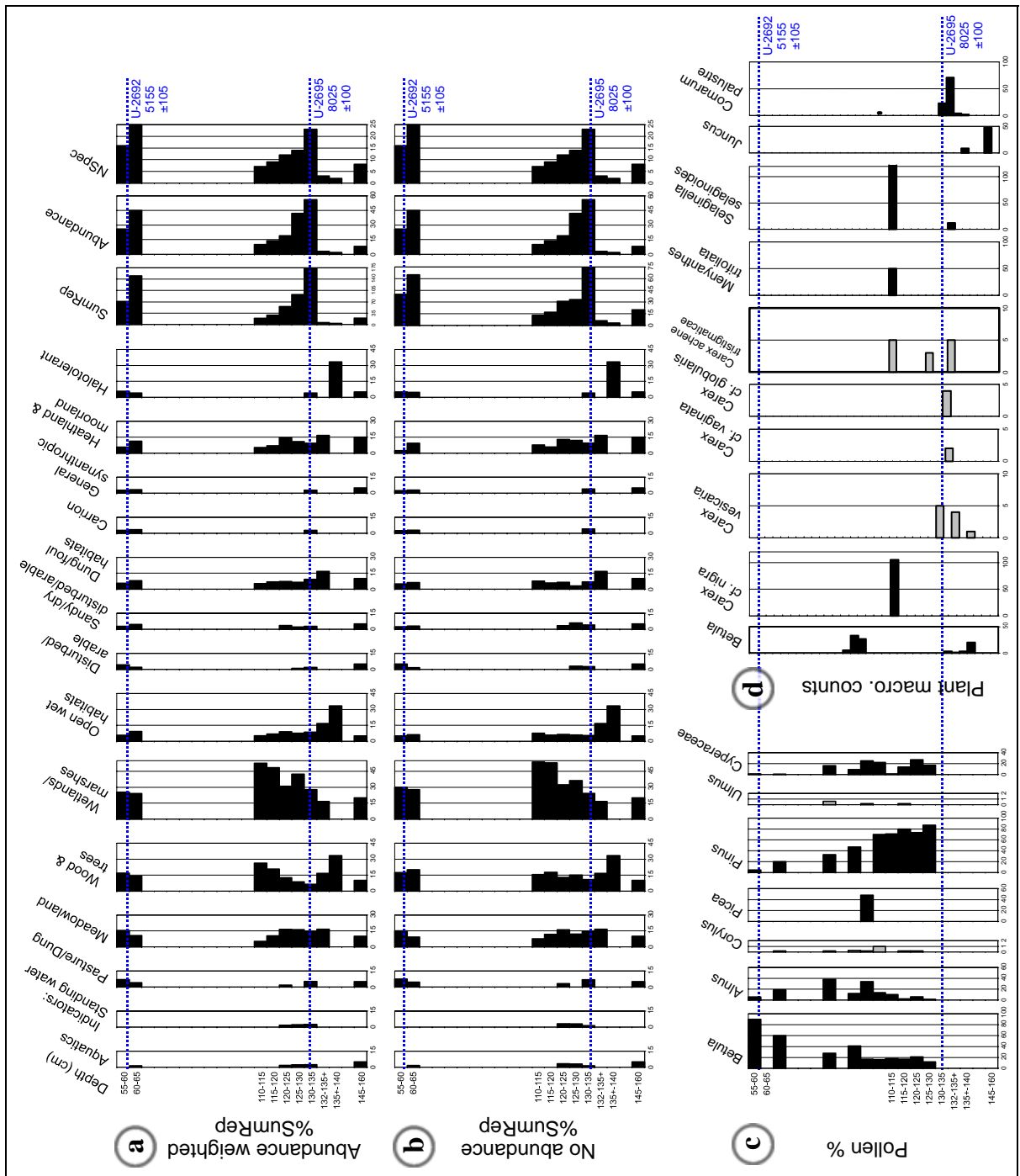


Figure 6.16. Bugs EcoFigs and other biological proxies for Hemavan. Standardized beetle reconstructions with (a) abundance weighting and (b) taxa only. Diagram (c) shows summary pollen percentage data, and (d) raw plant macrofossil counts. Radiocarbon dates are uncalibrated, and transferred by stratigraphic correlation.

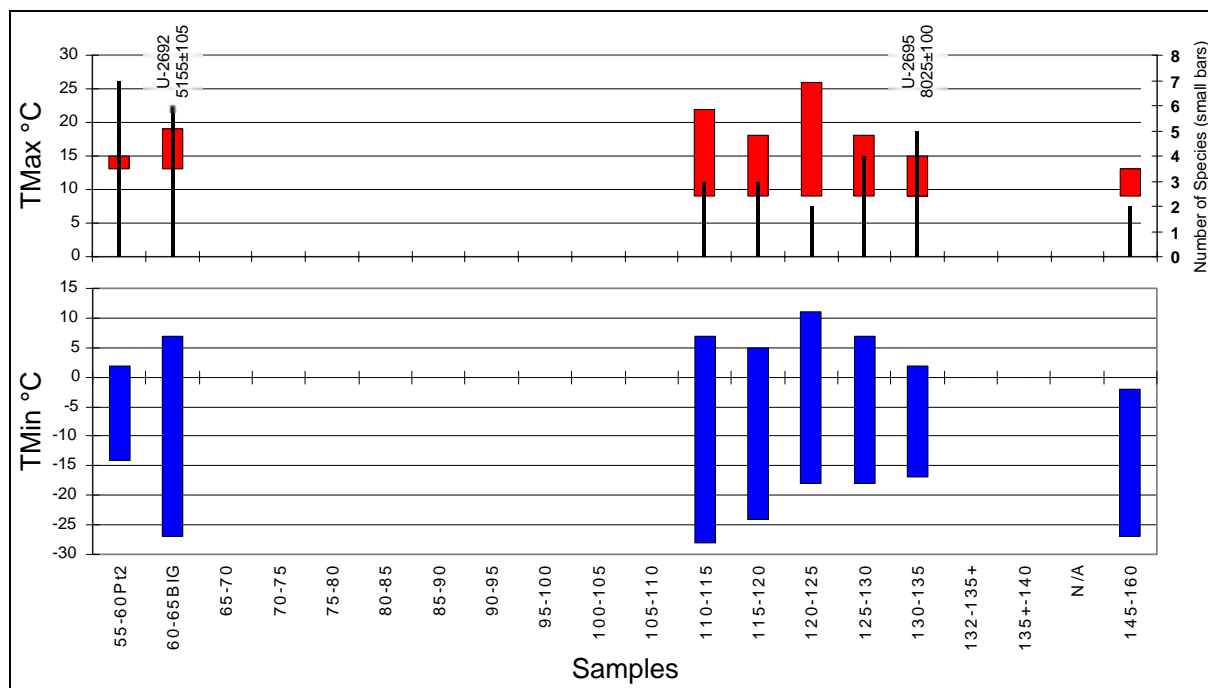


Figure 6.17. MCR reconstruction of palaeotemperatures from Hemavan. Uncalibrated ^{14}C dates are shown on the diagram; 1σ calibrated ranges U-2692: 5122-6184 BP and U-2695: 8599-9243 BP (Oxcal version 4, Bronk Ramsey, 1995). Present day temperatures: TMax: 8 to 10°C; TMin: -15 to -12°C (SMHI, 2005)

6.3.4 Discussion

As is often the case where sample columns cross sediment boundaries, there is considerable difference in the species richness and total abundance between samples. This necessitates the standardization of EcoCode counts to compensate for these variations, so that samples with higher diversity do not automatically gain stronger habitat indications. Where the sediment type is the same, however, it could be argued that the insect diversity could be a proxy for bog growth rate – a fast growing bog accumulating less individuals per unit time than a slow growing one. This assumes a constant rate of ‘capture’, which is difficult to prove without independent evidence from another deposit in the area.

The beetle faunas appear to depict an initial picture of a lake overgrowing into a bog. Unfortunately the number of species (2) and individuals (2) in samples 135+-140 and 132-135+ are too low for this to be realistically assumed. There is, in fact, very little indication of open water in the sequence, and the five most prominent habitat signals (*Meadowland*, *Wood & trees*, *Wetlands/marshes*, *Open wet habitats* and *Heathland & moorland*) are those that would be expected from a peat bog, with some small pools and wet mud on the surface, surrounded by woodland. There is strong evidence for the presence of standing water, in the form of the Bugs EcoCode indicator species *Hydroporus memnonius* Nic. (Nilsson & Holmen, 1995). The *Dung/foul* habitats implied in all samples, are almost all the result of generic level identifications of Staphylinid beetles of the genera *Lathrobium*, *Omalium*, *Philonthus*, along with *Cercyon* and *Aleocharinae*. The only individual representing this habitat that could be identified to species was the predatory rove beetle *Tachinus rufipes* (L.), which is as much at home in dung or decaying organic matter as in wetter grasslands and bogs (Andersen *et al.*, 1990; Koch, 1989). By far the dominant habitat indicated is that of *Wetland/marshes*, which would be expected for a peat bog. If one disregards samples 135+-140 and 132-135+ due to their low numbers, one can also see an indication of a gradual increase of woodland and trees around the bog.

There appears to be contradictory evidence when comparing the BugStats output with the pollen percentage data. The pollen data suggest a woodland landscape, initially dominated by pine just after 8 000 ^{14}C BP, gradually being replaced by birch, which totally dominates the pollen flora at 5 155 ^{14}C BP. This may reflect differing areas of representation of the insect fauna and the pollen

flora. Pollen is more easily wind transported than insects, with pine pollen being the extreme example. Many insects are also more inclined to directed travel, in that they will actively search for favourable habitat, rather than being passively borne on the wind. This increases the probability that an insect found in a trap in a particular environment will, by its own volition, have intended to be there and thus represent the environment. With this reasoning one can suggest, in this context, using the pollen to paint a picture of the wider surroundings, and the beetles and plant macrofossils to reconstruct the immediate bog surface, and to a lesser extent the immediate surroundings.

The plant macrofossil data shows early indications of bog/wetland growth in the form of *Carex* spp., *Comarum palustre* L., *Menyanthes trifoliata* L. and *Selaginella selaginoides* (L.). Birch is indicated, and most probably growing at the edges of the bog. These data support the local environment described by the insect data, and complement the more regional signal from the pollen data. Treeline megafossil evidence from the region (Kullman, 1992), although difficult to interpret in terms of populations, are broadly supported by the pollen evidence of Engelmark (1996) if one assumes that the conditions for a heightened species treeline also lead to a relative increase in the pollen abundance of the species at lower altitudes. More high altitude beetle studies are required to assess the ability of beetle faunas to reflect changes in treeline vegetation.

6.3.5 Evidence for early Holocene climate change

The lowest insect sample (145-160cm, >9 243 Cal. BP) may suggest a slightly colder thermal regime than the higher samples, somewhat similar to the present day temperatures at Hemavan. Although the range of possible temperatures overlaps the majority of the samples (the exceptions being the highest processed samples from 55-65 cm), the fact that the upper limits of both TMax and TMin are lower than in any other samples may be significant. This result is somewhat steered by the presence of *Acidota quadrata* (Zett.) which has a particularly narrow TMax span of 9 to 13°C. Using the 'Predict' function in BugsMCR, it can be seen that the only other species in the MCR database that could fit within this range is the cold tolerant but more hygrophilous *Elaphrus lapponicus* Gyll. Both species are known from Late Glacial samples from sites in the UK (e.g. West Bromwich (Osborne, 1980); Red Moss (Ashworth, 1972)), Denmark (Nørre Lyngby, (Coope & Böcher, 2000)) and Sweden (e.g. Körslättamossen (Hammarlund & Lemdahl, 1994); Björkeröds Mosse (Lemdahl, 1988))^{xii}. The present day geographical distributions of both species in Sweden are very similar, and show a distinct northern preference (Gustafsson, 2005), suggesting that the reason for the presence of only one of them in the sample may not be explainable in terms of distributional factors. In which case, one could look to ecological factors and taphonomy (including chance) to explain the absence of *E. lapponicus* from a sample that it theoretically could populate. It could be that the Early Holocene climate of Hemavan was too dry for this species. Although this is a potential explanation, the small number of individuals in the sample (8) give more weight to an explanation in terms of a low probability of incorporation into the sample, and the suggestion that a much larger sample would be needed to assess the hypothesis.

Table 6.9. Thermal limits for species in Hemavan bottom sample 145-160cm. See Chapter 5 for explanation of variables.

Species	TMaxLo	TMaxHi	TMinLo	TMinHi	TRangeLo	TRangeHi
<i>Acidota quadrata</i> (Zett.)	9	13	-30	-2	15	39
<i>Tachinus signatus</i> Grav.	9	27	-27	15	7	36

The only other MCR species present in the sample, *Tachinus signatus* Grav. is a cold tolerant but significantly more eurythermal species (Table 6.9), currently known from the whole of Sweden, which may be living towards the lower limits of its thermal tolerance in this sample. In contrast to *A. quadrata* there are 235 species in the database which could survive within the TMax limits of *T.*

^{xii} This is not an exhaustive list, consult BugsCEP for further sites.

signatus. Koch (1989-92) classifies this species^{xiii} as ubiquitous (Habitat range: Ub) and saprophilous (Food type: sa; in decaying materials), and Good & Giller (1991) record it as a predator on the larvae of other Coleoptera, Diptera and Collembola. These are characteristics which would give the species a good chance of survival in a recently deglaciated landscape.

Although the samples from 110-135 cm provide temperature reconstructions which overlap those from the basal clay sample (145-160 cm), the samples at 55-60 cm and 60-65 cm suggest mean July temperatures of 13-15°C and 13-19°C, the lower limits of which represent a regime some 3-5°C warmer than the present day. Dated to 5122-6184 Cal. BP, this is only slightly later than the expected early Holocene warm period suggested by other proxies for northern Sweden (e.g. Bigler *et al.*, 2002). Analysis of the remaining samples is necessary to investigate as to whether this is an anomaly, or part of a pattern reflected in other mid-Holocene samples from Hemavan, such as the development of more oceanic conditions. There is currently insufficient evidence in the beetles to assess the evidence for climatic change presented by the megafossil data of Kullman (1992).

6.3.6 Conclusions

The environmental reconstruction from Hemavan give a picture of Early Holocene bog growth which is compatible with current understanding of the post-Glacial development of the region, and indicated the complementary nature of palaeontomological, plant macrofossil and palynological evidence. The reconstruction temperatures are compatible with those derived from other proxies from northern Sweden (e.g. Rosén *et al.*, 2001), although much of the Hemavan sequence is yet to be analysed. More samples will need to be examined in order to produce a fuller picture of Mid-Holocene environmental and climate change in the region, and deposits from other bogs and lakes in the area should be examined to enhance the understanding of a poorly researched region.

6.4 Fossil Case Study: Two ‘Wells’ at the Archaeological Site Lockarp 7B, Sweden.

Own data contribution: fossil insect data

Site reference: Eliasson & Kishonti (2003)

6.4.1 Aims and introduction

The aim of this study was to examine the beetle fauna of an archaeological feature described as a well through the BugsCEP software, and create a reconstruction of the environment represented by the species present. The fossil insects from archaeological well deposits of different ages have been studied by a number of authors (Late Neolithic: Schelvis, 1989; Bronze Age: Masefield, 2003; Osborne 1969, 1989; Iron Age: Antipina *et al.*, 1991; Hellqvist, 1999; Hellqvist & Lemdahl, 1996), and especially from Roman well deposits (Alvey, 1976; Buckland, 1980 & 1986; Coope & Osborne, 1968; Girling, 1989; Hakbijl *et al.*, 1989; Kenward *et al.*, 1986; Lentacker *et al.*, 1992; Mertens *et al.*, 1986; Osborne, 1975; Simpson, 2001; Sudell, 1990; and others). Wells have been compared with giant pitfall traps (Buckland, 2000; Hellqvist, 1997), although their location will be critical to their trap effectiveness. Covered or indoor wells, for example will most probably produce much poorer faunas than open or outdoor ones.

The interpretation of wells is generally difficult, and especially so where the archaeological interpretation is problematic. There is a tendency, in Swedish archaeology at least, to interpret almost every significant excavated filled depression as a well, even before palaeoecological analysis have been undertaken. Such an *a priori* hypothesis is only acceptable as long as it does not interfere with

^{xiii} See Chapter 4 for more information on Koch’s ecology codes.

interpretation, especially if the insects and plant macrofossils provide evidence to the contrary. So called ‘wells’ equally often turn out to have been dew ponds, animal drinking troughs or cisterns, where the water quality was almost certainly unfit for human consumption on the evidence of the insects (e.g. Jaques *et al.*, 2000), although by the time accumulation had begun in the feature it may have ceased to be a water source.

The samples analyzed here form part of a large archaeological project connected to the construction of a major new roadway system around Malmö, Sweden (Eliasson & Kishonti, 2003). A number of other analyses, including plant macrofossils, snails and soil chemistry were undertaken on the samples by the Environmental Archaeology Lab at Umeå, Sweden, summary reports for which can be found in Eliasson & Kishonti (2003). A synthesis of the environmental evidence will be available in a forthcoming publication (Engelmark & Linderholm, *in prep.*).

6.4.2 Methods

The two sets of samples interpreted here were submitted to the Environmental Archaeology Lab in Umeå as coming from two wells, numbered 14495 and 26551. Both features were sampled with kubiena box columns from the excavated profiles, and the stratigraphy of both features is shown in Figure 6.18 and Figure 6.19. Although sample size was small it was possible to retrieve insect remains from the lower part of the column, where preservation was sufficient for identification. These samples were washed through a 300 µm sieve and examined under a binocular microscope. Insect remains were extracted and identified by the author of this thesis, under the guidance of Paul Buckland, with reference to identification keys and modern reference collections. Environmental interpretation was undertaken with the help of the BugsCEP software, and reconstruction diagrams were produced using the BugStats component (Figure 6.20).

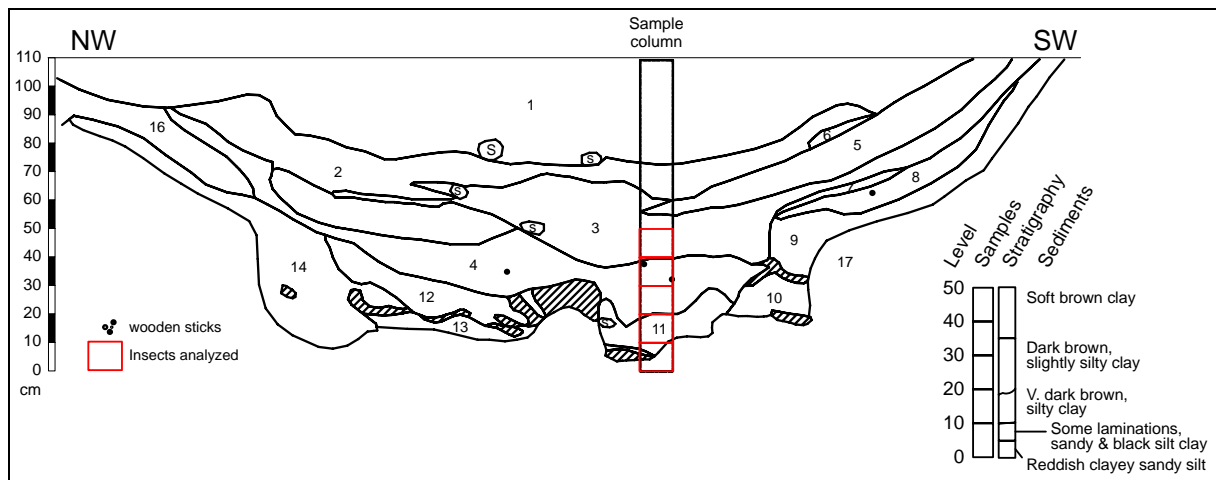


Figure 6.18. Stratigraphy and sample locations in feature 14495, Lockarp 7B. Stratigraphy provided by Johan Linderholm; archaeological sediment descriptions can be found in Eliasson & Kishonti, (2003).

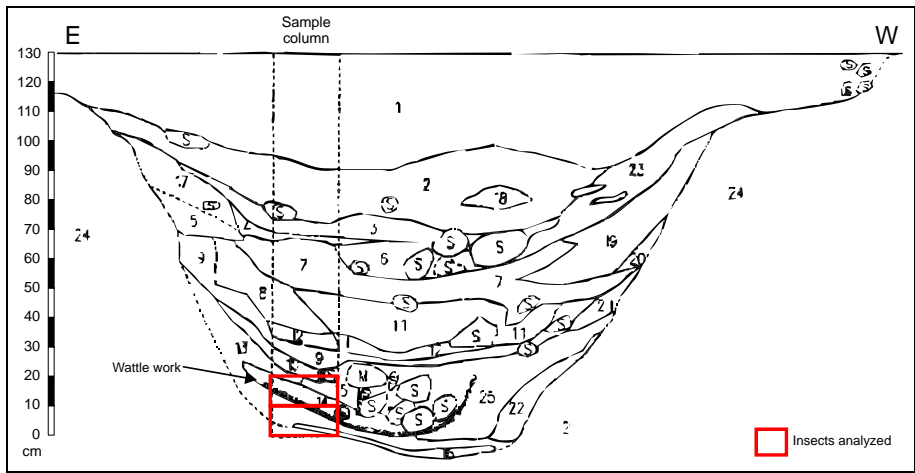


Figure 6.19. Stratigraphy of feature 26551, Lockarp 7B. Insects were only preserved in the bottom 20 cm of the sequence. Stratigraphy provided by Johan Linderholm; archaeological sediment descriptions can be found in Eliasson & Kishonti, (2003).

6.4.3 Results and discussion

A total of 35 taxa were identified from the five samples from feature 14495 (Table 6.10), and 40 from feature 26551's two samples (Table 6.11). Species abundances were low, with a maximum MNI of 4 for any taxon. Species richness was generally low in all samples, with sample 10-20 from feature 26551 producing the highest number of taxa. Small sample size and variable preservation are responsible for these low numbers. Only %SumRep outputs were produced, as the use of absolute abundances where the numbers are so low would produce the appearance of grossly exaggerated signals from the more plentiful taxa in a raw abundance weighted diagram. A raw taxa (no abundance) diagram could also be misleading in that the species richness of the samples varies considerably.

Table 6.10. Beetle species from Lockarp 7B feature 14495.

Taxa	S0-10	S10-20	S20-30	S30-40	S40-50
<i>Dyschirius</i> sp.		1			
<i>Bembidion properans</i> Steph.				1	
<i>Pterostichus vernalis</i> (Panz.)				1	
<i>Hydraena</i> sp.				1	
<i>Helophorus brevipalpis</i> Bed.			3	1	
<i>Helophorus brevipalpis</i> grp.			4		
<i>Laccobius</i> sp.				1	
<i>Silpha</i> sp.					1
<i>Anotylus rugosus</i> (F.)				1	
<i>Anotylus nitidulus</i> (Grav.)			1		
<i>Platystethus nodifrons</i> (Mann.)	1	2	1		
<i>Platystethus nitens</i> (Sahl.)				1	
<i>Stenus</i> sp.		1			
<i>Xantholinus</i> sp.		1			
<i>Dryops</i> sp.			1		
<i>Geotrupes stercorarius</i> (L.)			1		
<i>Onthophagus joannae</i> Golj.				1	
<i>Oxyomus sylvestris</i> (Scop.)					1
<i>Aphodius sticticus</i> (Panz.)		1			
<i>Aphodius obliteratus</i> Panz.				3	
<i>Aphodius contaminatus</i> (Hbst.)	1				
<i>Aphodius sphacelatus</i> (Panz.)			2		
<i>Aphodius prodromus</i> (Brahm)			2		
<i>Aphodius foetidus</i> (Hbst.)				1	
<i>Aphodius granarius</i> (L.)		2	2		
<i>Aphodius</i> sp.	1		2	1	
<i>Aphodius</i> spp.				3	2
<i>Longitarsus</i> sp.	1		2		
<i>Ochrosis ventralis</i> (Ill.)	1				
<i>Chaetocnema</i> sp.		1			
<i>Psylliodes affinis</i> (Payk.)		2	2		
<i>Psylliodes</i> sp.				1	1
<i>Sitona</i> sp.			1		
<i>Dorytomus</i> sp.			1		
<i>Ceutorhynchus</i> sp.			1		

Table 6.11. Beetle species from Lockarp 7B feature 26551.

Taxa	S0-10	S10-20
<i>Carabus violaceus</i> L.	1	1
<i>Bembidion properans</i> (Steph.)		2
<i>Pterostichus strenuus</i> (Panz.)		1
<i>Pterostichus niger</i> (Schall.)		1
<i>Calathus fuscipes</i> (Goeze)	1	1
<i>Hydroporus</i> spp.		3
<i>Ochthebius</i> sp.		1
<i>Sphaeridium bipustulatum</i> F.		1
<i>Sphaeridium lunatum</i> F.	1	2
<i>Cercyon melanocephalus</i> (L.)		1
<i>Cercyon</i> sp.	1	
<i>Megasternum obscurum</i> (Marsham)		2
<i>Manda mandibularis</i> (Gyll.)		1
<i>Aploderus caelatus</i> (Grav.)		1
<i>Aploderus</i> sp.	1	
<i>Anotylus rugosus</i> (F.)		2
<i>Anotylus nitidulus</i> (Grav.)		2
<i>Platystethus cornutus</i> (Grav.)		3
<i>Platystethus nodifrons</i> Mann.	1	2
<i>Platystethus nitens</i> (Sahl.)	1	4
<i>Stenus</i> sp.		1
<i>Philonthus</i> sp.	1	1
<i>Quedius</i> sp.		1
<i>Aleocharinae</i> indet.	1	3
<i>Athous haemorrhoidalis</i> (F.)		1
<i>Dryops</i> spp.		4
<i>Enicmus</i> sp.		1
<i>Corticaria</i> sp.	1	2
<i>Scymnus frontalis</i> (F.)	1	
<i>Scymnus</i> (s.l.) sp.		1
<i>Anthicus</i> (s.l.) sp.		1
<i>Aphodius erraticus</i> (L.)		1
<i>Aphodius rufipes</i> (L.)		1
<i>Aphodius contaminatus</i> (Hbst.)		2
<i>Aphodius foetidus</i> (Hbst.)		1
<i>Aphodius</i> sp.		1
<i>Aphodius</i> spp.		2
<i>Phyllotreta nemorum</i> (L.)		1
<i>Aphthona</i> sp.	1	
<i>Longitarsus</i> sp.	1	
<i>Longitarsus</i> spp.		3
<i>Chaetocnema concinna</i> (Marsham)	2	
<i>Psylliodes affinis</i> (Payk.)		1
<i>Psylliodes</i> sp.	1	
<i>Notaris acridulus</i> (L.)		1
<i>Tychius</i> sp.		1
<i>Salda</i> sp.		1

6.4.3.1 Feature 14495

The low abundances are reflected in the little difference between the abundance weighted (Figure 6.20a) and no abundance (Figure 6.20b) outputs from feature 14495. A slightly higher aquatic signal was derived from the abundance weighted reconstruction of the 20-30 cm level, along with a few minor variations in the other signals, but the broad patterns are the same. The most immediately obvious feature is the *Aquatics* signal between 20-40 cm, which is mirrored with a drop in *Meadowland* habitats and a slight indication of *Open wet habitats*. In a set of samples with such low diversity there is a slight chance that this is an artefact, as the surrounding samples have fewer species and individuals and so probability dictates that they are less likely to reflect as many environments as those with higher numbers. Since this feature has been archaeologically interpreted as a well, however, it is reasonable to assume some truth in the signal. It could be that these two samples represent the only true well phase in the sequence, and that the other samples represent construction and collapse phases. Alternatively, and this could be supported by the increased diversity, these two samples represent an open well phase where more beetles accumulated in the well.

There is a relatively strong dung signal throughout the sequence, especially in the bottom 40 cm where it is represented by indicator species. These species, which include *Geotrupes stercorarius* (L.), *Aphodius contaminatus* (Hbst.), *Aphodius sphaelatus* (Panz.), *Aphodius prodromus* (Brahm) and *Onthophagus joannae* Goljan, are highly dependant on animal dung for their survival and reproduction, and thus their presence in samples is as near to 100 % an indication as science can provide^{xiv} (Geoffrey Lemdahl, *pers. comms*). Some of them are also specific to particular substrate materials, and can tell more about the area around the well than just the fact that there were very probably large herbivores there. *O. joannae* is a good indication of sandy or calcareous soils (Duff, 1993; Jessop, 1986).

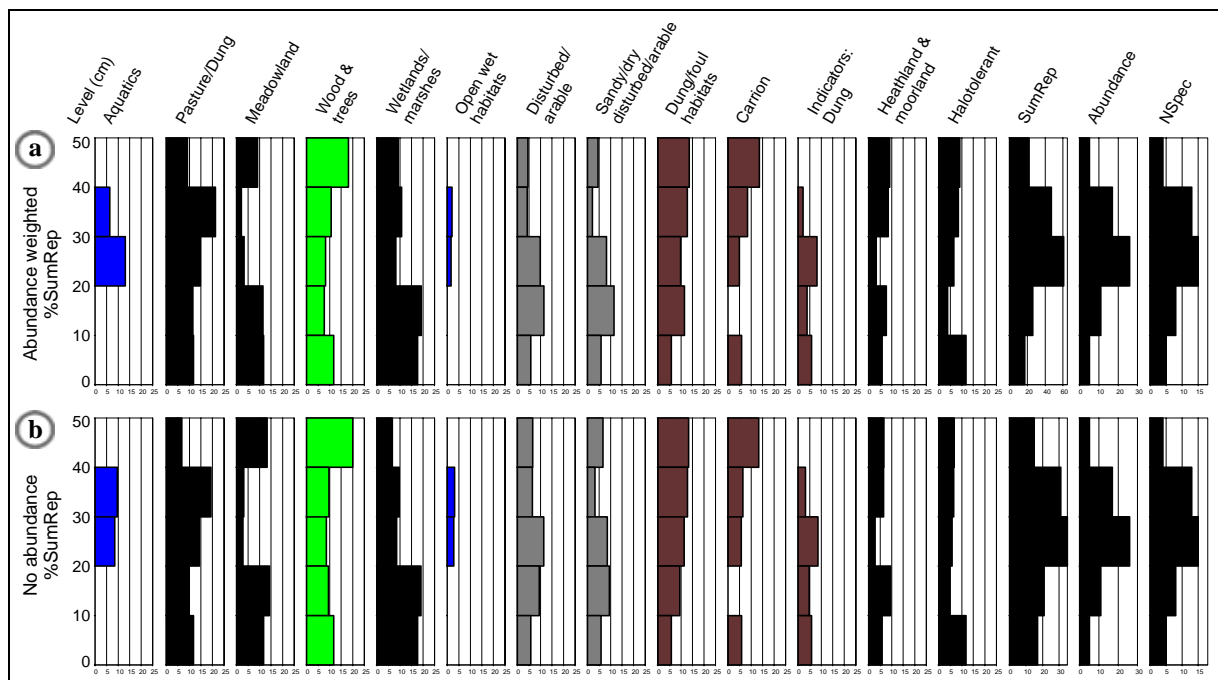


Figure 6.20. BugStats output for Lockarp 7B feature 14495. Both diagrams are standardized, showing (a) abundance weighted and (b) taxa only results.

^{xiv} Gunnar Gustavsson provided help with the indicator species list. *A. sphaelatus* may not necessarily belong in the indicator class (as suggested by Koch 1989 and others), as a number of sources indicate that it is often found in decaying vegetation as well as dung (Duff, 2003; Jessop, 1986; Landin, 1961; Machatschke 1969). Geographical variation in stenotopy creates classification problems (as described in Chapter 4).

The *Dung/foul habitats* class caters for the fact that a number of species are equally happy in other decaying matter as well as dung, and the *Carrion* category has some overlap with this. These are good indications of what would be considered as a farmyard environment today, but in fact, would probably be fairly typical of any dwelling site or settlement before the early modern period, when a clearer separation between human and animal occupation areas began. There appears to be a steady increase in indications of pasture (*Pasture/Dung*) from the bottom of the sequence, provided by an increasing proportion of *Aphodius* species in the samples, although identifying such trends through samples with so few individuals is dangerous.

The *Halotolerant* signal, which is more or less constant after an initial peak, is entirely the result of generic level identifications of the ground beetles *Dyschirius* sp., dung beetles *Aphodius* sp(p)., and flea beetles *Longitarsus* sp.. In this case it should probably be ignored, as there are few too individuals involved for the signal to be probable.

6.4.3.2 Feature 26551

As can be seen from Figure 6.21, the two samples from this feature have number of similarities with samples S10-20 and S20-30 from feature 14495 (Figure 6.20), which could be an indication that they both represent the transition from primary sediments to water deposited layers. The lower sample has no indication of *Aquatics*, but a reasonable indication of *Wetlands/marshes*, suggesting some wetland vegetated areas in the surroundings, but no open water. Sample S10-20, on the other hand, shows a definite water signal in the presence of water beetles of the genera *Hydroporus* and *Ochthebius*. There is a *Halotolerant* signal in both samples, suggesting saline environments. Closer examination of the fauna reveals, as above, that this is the result of generic and family level identifications, and it is therefore not a reliable indicator in this case. The calcareous soil indicated in feature 14495 (by *Onthophagus joannae* Goljan) is further supported by the presence of the large ground beetle *Carabus violaceus* L., the lady bird *Scymnus frontalis* (F.), and to a lesser extent the ground beetle *Calathus fuscipes* (Goeze). This is not surprising for this location in Skåne due to the local geology.

Perhaps the most interesting difference between this feature and 14495 is the increased presence of *General synanthropic* species – those that are generally associated with environments strongly influenced by human actions. This habitat group overlaps to some extent with *Mould beetles*, and the taxa in this case suggest an amount of unspecified decaying material in the surroundings. Many of these species, including members of the *Enicmus* and *Corticaria* genera found in the samples, are common indoor guests and, when found in larger numbers can indicate stored foodstuffs or mouldy hay or timber. Low abundances, and generic identifications here, however, prevent further conclusions from being drawn. The Elaterid *Athous haemorrhoidalis* (F.) is entirely responsible for the *Dry dead wood* signal, which may not be entirely correct. Whilst the adults appear in woodland, the larvae live on the roots of grasses.

A general picture of a messy, inhabited dwelling site is presented by sample S10-20, and to a lesser extent by sample 0-10. Only the upper sample shows any real indication of the presence of water, and the sediments may have been deposited in a well or animal watering hole.

The various BugStats outputs for feature 26551 have useful illustrative value, in that they clearly demonstrate the need for standardizing reconstruction values to enable proper inter-sample comparisons to be made. The difference between the standardized plots is minimal – the low abundances leading to small multipliers for the taxa, which do not give any significant change in the overall shape of the diagrams. The raw diagrams also appear reasonably similar to each other, if one ignores the x-axis scales, which obviously are much larger on the abundance weighted figure. However, when one compares the raw diagrams with the standardized diagrams, a considerable difference can be seen. The extreme difference between the two samples, in terms of species richness and abundance (Table 6.12), are reflected in the raw diagrams by sample 10-20 having higher indications of environments represented. Standardization allows samples to be compared by proportioning the sample diagram components to the total sum of environmental representations for

the sample. In this case we can see that the relative extent to which some environments are represented in each sample is reversed between the figures. Evidently the high abundances in sample 10-20 bias the reconstruction.

Table 6.12. Summary counts and sums for the Lockarp 7B ‘wells’.

Feature 14495						Feature 26551		
Sample	S0-10	S10-20	S20-30	S30-40	S40-50	S0-10	S10-20	
No. of taxa	5	8	15	13	4	15	40	
Total abund.	5	11	26	17	5	16	63	
Max abund.	1	2	4	3	2	2	4	

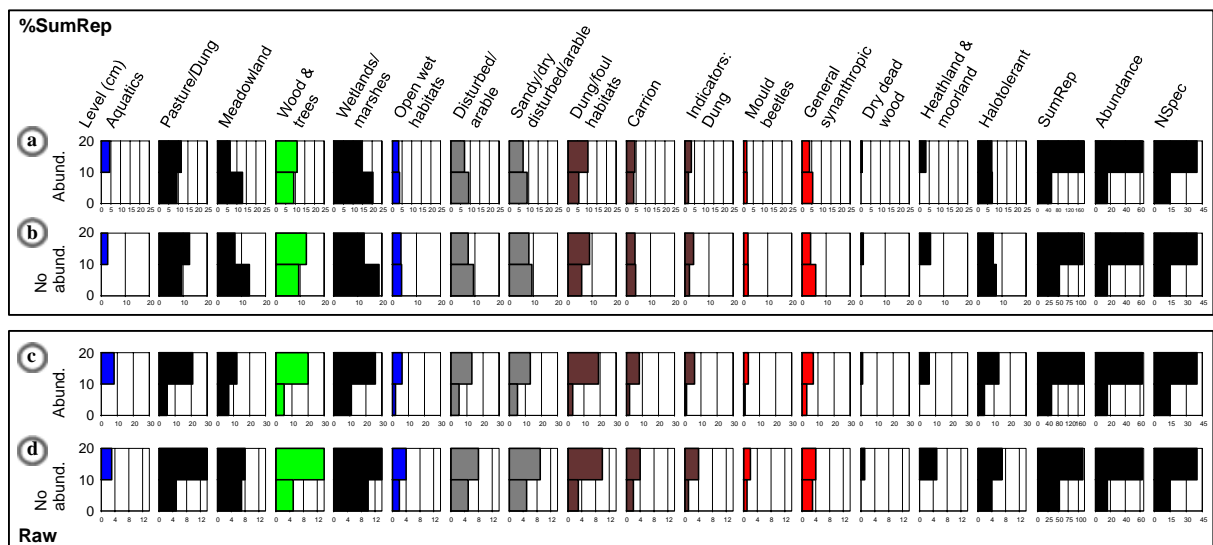


Figure 6.21. BugStats output for Lockarp 7B feature 26551, illustrating the importance of sample standardization (see text). Diagrams (a) and (a) show standardized abundance weighted and taxa only results respectively; diagrams (c) and (d) show raw counts, abundance weighted and taxa only respectively.

6.4.4 Conclusions

From the beetle evidence, Feature 14495 could very probably be a well in a settlement context, with samples 20-30 and 30-40 representing an open well in active surroundings. With such small numbers, however, it is not possible to tell whether the water would have been fit for human consumption. From the dung indications, it is most likely that animals were present in the surrounding throughout the sequence. The samples below and above this most likely represent construction and collapse phases respectively, and contain little to indicate a well or water. Although there were only a few individuals found in the latter samples, they give some indication of the catchment environment. These signals, however, are difficult to interpret due to the possibility of being secondary deposits of uncertain origin. The upper sample of Feature 26551 may represent a water trough or well in a settlement environment, with animals and decaying organic matter in close proximity.

The Lockarp results present a considerably more complex environmental picture when compared to the results from the analysis of the trap data from Koivula *et al.* (2003) (section 6.1). At Lockarp we see indications of 13 and 16 of the 22 classes in the Bugs EcoCode system, whereas the Finnish trap data indicated only 8. The most probable explanation for this is that the Koivula *et al.* data represent only one year, whereas the Lockarp samples represent an unknown, but probable multiple year span in a much more diverse environment. The longer a trap remains open, the greater the probability of rarer species being caught in it, just as the probability of surveying all species in an environment increases

with the length of survey. This relationship has been studied in detail by ecologists (Krebs, 1989), and is often seen to be log/normal – the number of new species found/trapped decreases with time. The second possible explanation is that the environment of Lockarp was indeed more complex (or patchy) than the forest-farmland transects. Urban areas/settlements provide a greater variety of microhabitats than natural or rural areas, and so will result in a greater diversity of environmental signals being given from samples.

The interpretation of fossil insect faunas requires a careful balance in the understanding of the probability of environments represented in the faunas, the nature and origin of the deposits, taphonomic processes, and ecological possibilities.

6.5 Preliminary Fossil Results from Lake Njulla, Abisko, Sweden.

6.5.1 Introduction

An attempt was made to include combined fossil insect data from the correlated contents of 14 cores from lake Njulla, near Abisko, Sweden (68°22'N, 18°42'E, 999 m a.s.l.) (Figure 6.22). The samples were provided by Petér Rosen of the Climate Impacts Research Centre, Abisko. Despite the assistance of Peter Rosén, Roger Engelmark and Geoffrey Lemdahl the processing and analysis of about 500 small samples did not prove feasible^{xv} in combination with the software development which is the primary focus of the thesis. Cores were correlated optically, using digital photographs, greyscale imaging and simple manipulations to identify comparable features. All cores were taken from a small region (c. 5 m in diameter) in the area of earlier investigations (Barnekow, 1999a; Kullman, 1999; Rosén, *pers. comm.*). Magnetic susceptibility (MS) measurements were taken along cores 8 and 11 which backed up the visual correlation, although the accuracy of MS to correlate the cores was deemed insufficient for it to be used as the primary method.

The intention was to obtain beetle data to complement the existing proxy analyses from the same site (e.g. Barnekow, 1999a; Bigler *et al.*, 2003; Rosén *et al.*, 2001), and compare the temperature and environment reconstructions. Particular attention would be paid to the presence of early indications of trees, early Holocene megafossils of which (*Pinus*, *Betula* and *Alnus*) have been found at the site by Kullman (1999). Analysis of the cores is still underway, and although species numbers are too low in the individual cores to permit detailed analyses, the correlation will allow the results to be bulked up between cores.

^{xv} Even at the rapid rate of half a day per sample, this would take 250 days.

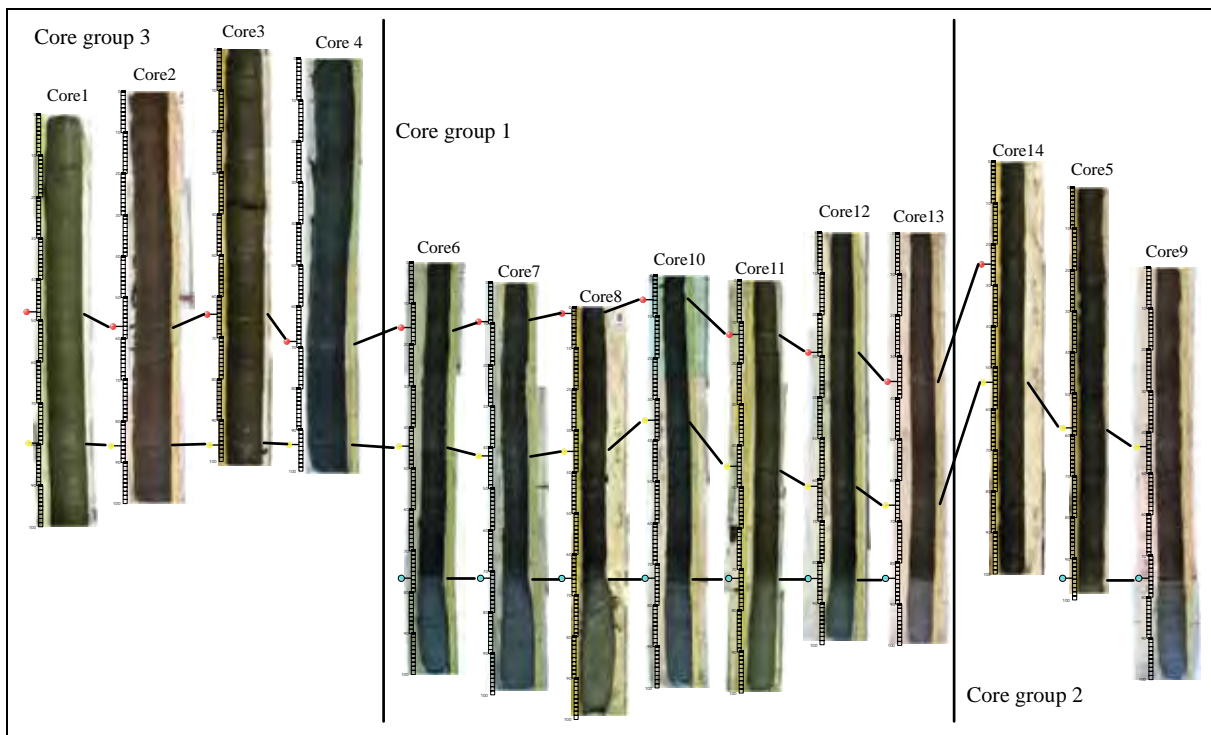


Figure 6.22. Correlation of cores from Lake Njulla, Abisko, Sweden. Lines indicate correlation horizons between the cores. Core group 3 was taken with a larger corer bore and was not able to penetrate as deep as the others. Core group 1 is considered as the most reliable set of cores.

6.5.2 Preliminary results

Partial results are available from cores 6 and 8^{xvi}, although no attempt has yet been made to bulk up the numbers of individuals by core correlation. Approximate ages have been derived by comparison of the gyttja-clay boundary found in Njulla lake samples by Barnekow (1999b) and Bigler *et al.* (2003), and suggest that analysed material dates from between 3 880 ¹⁴C BP and 8 200 ¹⁴C BP. The present day (1960-1990) mean July temperature of the area is c. 8.1°C Barnekow (1999b), and the preliminary reconstruction shows the majority of Early-mid Holocene temperatures to have been at least, and probably more than, 1°C warmer.

The gyttja-clay interface, dated to 8 200 ¹⁴C BP by Barnekow (1999b), of Core 8 appears to suggest a slightly colder winter temperature, when compared to the samples from the levels above, as indicated by the cold tolerant *Pycnoglypta lurida* (Gyll.). This species is known from a number of Lateglacial and mid-Devensian sites in the UK (see the BugsCEP records for a list of sites). There may be a slight indication of an Early Holocene, between c. 6 960 ¹⁴C BP and c. 8 200 ¹⁴C BP, warmer than the rest of the sequence, although the data are too sparse for any distinct conclusions to be drawn. Core 6 displays a similar Early Holocene thermal picture to Core 8. There are also important climatic indicator species present, such as the continental, cold tolerant, *Mannerheimia arctica* (Er.), which have no reference data in the current MCR dataset.

The preliminary habitat reconstruction results suggest a heathland environment, with more or less continuous indications of water nearby, as would be expected for samples from high altitude lake deposits. Interestingly, the only indications of woodland and trees occur during the aforementioned potentially warmer period, which supports the macrofossil evidence of Barnekow (1999b) and potentially Kullman (1999), although indicators of specific tree species have yet to be found. The evidence, however is only from three individuals of the genus *Quedius*, which are predatory and able to utilise under a broad range of habitats, and should be regarded as tenuous until more specimens are

^{xvi} Processed by Philip Buckland and Peter Rosén, and identified by Geoffrey Lemdahl.

identified. There are also potential dung indicators, such as *Tachinus laticollis* Grav. and *T. rufipes* (L.) present during this period, although these particular species are quite eurytopic. The only indication of running water is to be found at the gyttja-clay interface, in the form of *Elmis aenea* (P. Müller), and the lack of such species in later deposits could potentially be an indication of the absence of flowing glacial melt-water which was present at the start of the sequence.

6.5.3 Conclusions

Bigler *et al.* (2003) showed that although July temperature reconstructions from pollen, diatoms and chironomids for lake Njulla provided consistent patterns for the last 6 000 years, the Early Holocene evidence was variable. Although individuals have been identified from only a few samples, the beetle evidence shows much potential for providing a terrestrial proxy for early Holocene climate change in the Abisko region, and thus complementing the existing interpretations. In addition, the potential of Coleoptera to indicate the presence of vegetation types difficult to identify in the pollen and plant macro records may help provide a more detailed picture of the evolution of the landscape surrounding the lake.

6.6 Modern Case Study: Pitfall Trap Data from the area of Gården under Sandet (GUS), Greenland.

Primary references: Buckland *et al.* (1998); Buckland (2000)

6.6.1 Aims and introduction

This small modern fauna was examined by this author in Buckland (2000) and with others in Buckland *et al.* (1998). It is briefly examined here to see how the application of Bugs EcoCodes affects the interpretation, and to test the investigative power of BugStats on a low diversity, low species richness modern species list. A comparison is made with principle component analysis (PCA) of the same fauna for illustration.

During the summer of 1995 a pitfall trapping and search exercise was conducted, by the author of this thesis, during the period of excavation of the Norse Farm site Gården under Sandet (GUS) in the area of the former Western Settlement of Norse Greenland (Buckland *et al.*, 1998). Vegetation zones were described onsite by Julie Ross (*pers. comm.*), and are summarised in Table 6.13. Information on the trapping setup and other results can be found in Buckland (2000).

Table 6.13. GUS modern, vegetation zone field descriptions.

Zone	Description
A	Upper floodplain, very low birch and sedges
B	Dwarf birch-rhododendron scrub (<i>Betula - Rhododendron</i>)
C	Bog – moss and coarse grasses
D	Willow dominated woodland with Labrador Tea under growth (<i>Salix - Ledum</i>)
E	Pondside marshland, sedge and moss dominated
F	Exposed hillside grassland

6.6.2 Results

The short species list is presented in Table 6.14. The zone sums are the total Coleoptera catches for all eight traps in a zone over a four week period from mid June to mid July.

Table 6.14. GUS modern species list, zone abundance sums for the four week collection period.

Species Name	Asum	Bsum	Csum	Dsum	Esum	Fsum
<i>Patrobis septentrionis</i> Dej.	0	0	1	0	0	0
<i>Hydroporus morio</i> Aubé	0	0	0	0	1	0
<i>Colymbetes dolabratus</i> (Payk.)	0	0	0	0	4	0
<i>Gyrinus opacus</i> Sahl.	0	0	0	0	1	0
<i>Mycetoporus nigrans</i> Maekl.	1	3	1	0	6	0
<i>Simplocaria metallica</i> (Strm.)	0	2	0	0	1	0
<i>Simplocaria elongata</i> Sahl.	0	1	0	0	0	0
<i>Byrrhus fasciatus</i> (Forst.)	2	2	1	0	0	1
<i>Nephus redtenbacheri</i> (Muls.)	0	0	0	0	0	2
<i>Coccinella transversoguttata</i> Fald.	4	1	0	1	0	0
<i>Otiorhynchus arcticus</i> (O. Fabricius)	1	1	1	1	0	1

6.6.3 Discussion

The zone sum data were imported into BugStats, and two output options chosen for presentation: non-weighted percentage, and abundance weighted raw, to illustrate the problem of interpreting small faunas. Figure 6.23 shows the results of the summed trapping and hunting data for each zone (A-F). What is immediately apparent is that Zone E, the pondside, is clearly visible by the graphs of aquatics and standing water indicators. The bog, Zone C, is also visible in terms of a wetland/marshland indication, but to a lesser extent. The other zones are not so easily identified from the graphs. The upper floodplain (A), dwarf scrub (B) and grassy hillside (F) show almost identical results, although the importance of abundance is clearly shown on the raw data diagram. This may indicate that these habitats are within the scope of variation for open, grassy landscapes in Greenland, or that the system is not sensitive enough to identify these environments with such small numbers of individuals. The two individuals found in Zone D, the (dwarf) willow woodland, vaguely indicate *Meadowland* and *Heathland & moorland*, but give no forest indication, being eurytopic, rather than woodland species.

These results suggest that low diversity/abundance data are, unsurprisingly, not ideal for environmental reconstruction. This indicates that the interpretation of small fossil assemblages should be undertaken with equal care, and that a qualitative indicator approach is likely to be more productive than semi-quantitative or statistical reconstruction. However, the fact that the bog and pond were correctly identified does prove that even small datasets may have some use, even if it is very biased towards very general habitat groups such as the aquatics.

As a comparison, the first two PCA components of the same data were examined with respect to species and samples (Figure 6.24 and Figure 6.25). Although PCA is not recommended for such small datasets, it is used here to illustrate the potential of BugStats for extracting information from these. BugStats failed to clearly illustrate more than the pondside and bog zones, and the PCA results are not particularly informative.

From Figure 6.25 we can see that Zone E (pondside) differentiates itself from the other zones on the first component axis. This could be explained in terms of the aquatic nature of the sample fauna, although the hypothesis of the first component representing a hydrological gradient is weakened by the position of Zone C (bog), the (qualitatively) next wettest site, at the middle of the spread of data. This could either be explained in terms of the particularly dry year, or in terms of the representativeness of the small fauna. On the second component axis of the species diagram (Figure 6.24) one can see the effect of the relative high abundance of *Mycetoporus nigrans*, a small rove beetle inhabiting moss and humus layers, which was new to Greenland with this work. Although the species is stenotopic to these environments, these themselves are quite widespread around the GUS site.

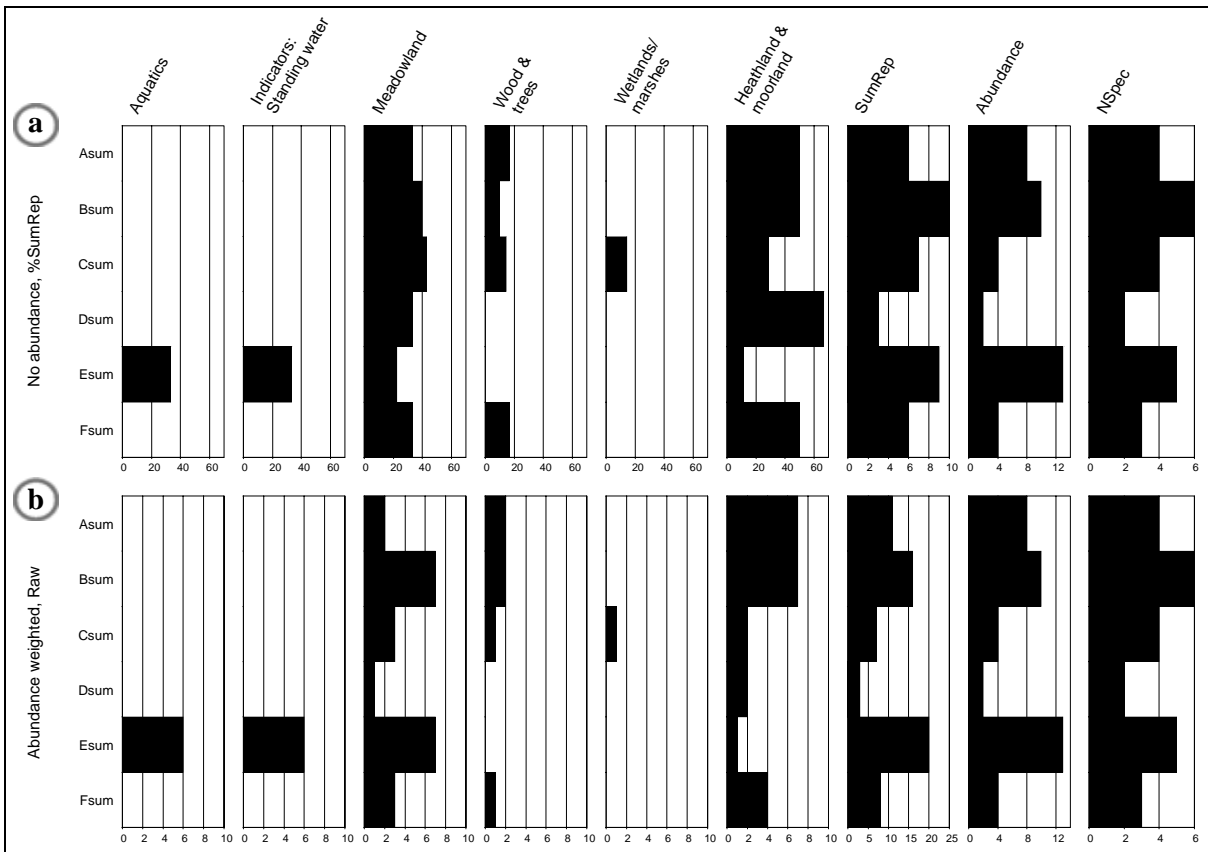


Figure 6.23. BugStats output for GUS modern data, showing (a) standardized, taxa only reconstruction, and (b) abundance weighted, raw counts based reconstruction.

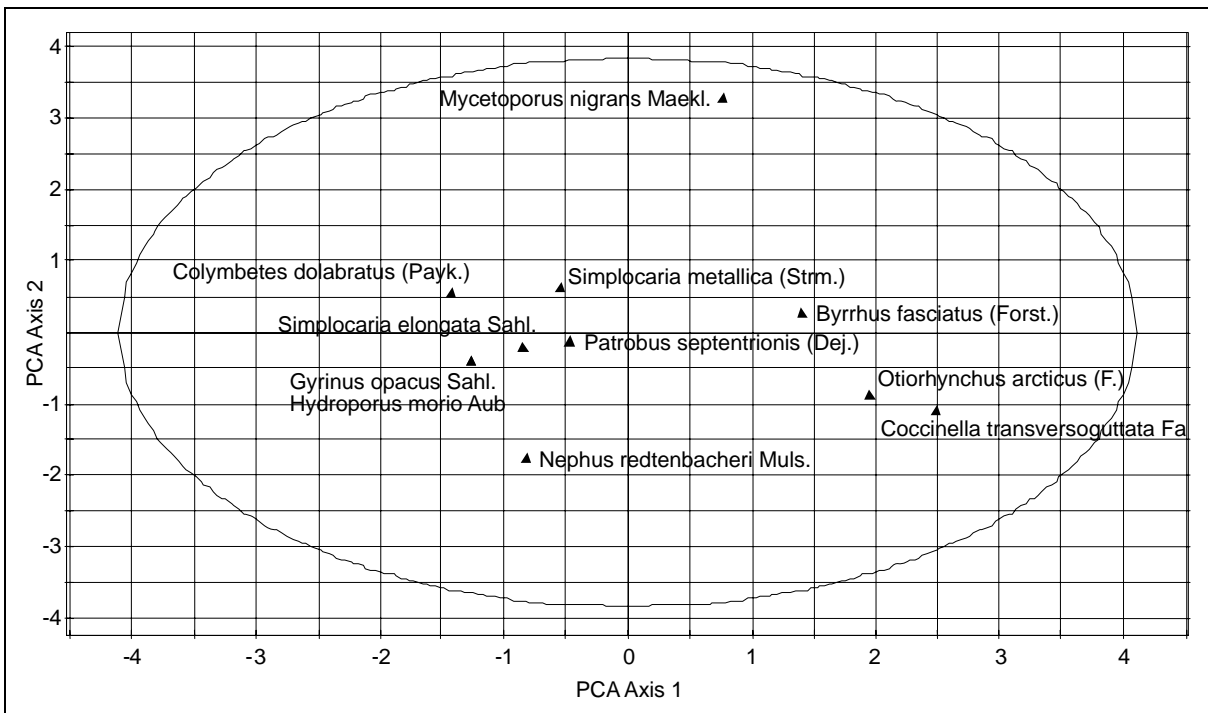


Figure 6.24. First two components of PCA on the GUS modern pitfall trap data, showing species names.

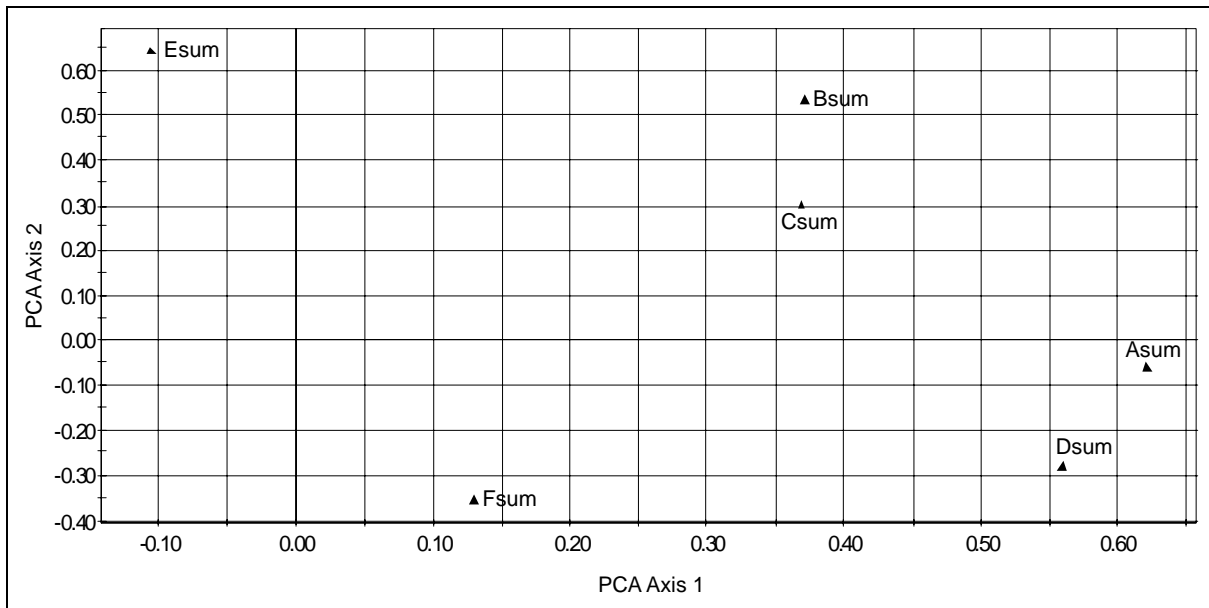


Figure 6.25. First two components of PCA on the GUS modern pitfall trap data, showing vegetation/sampling zones.

6.6.4 Conclusions

If one were to consider the fauna above as a fossil assemblage, one would be hard pressed to identify correctly more than the wetter zones E and possibly C. The presence of a number of water beetles in the former is enough indication of the presence of water close by. The relative dampness of the bog Zone C could be established from the examination of the fauna, but is not easily seen from the PCA. Neither the PCA grouping of Zones B and C, nor A and D, is easily explained in terms of the local environment. There is a risk that further attempts to explain such low abundance and diversity sites would lead to over-interpretation. In conclusion, without prior knowledge of the environment at the time of deposition, PCA is of little use on this type of small fauna. BugStats on the other hand was able to identify specific habitat components, due to its use of internal habitat reference data for each taxon. It may therefore be a more useful tool when undertaking habitat reconstruction from Quaternary deposit than ordination techniques.

6.7 Fossil Case Study: Climate and Environmental Change in Europe over the Past 20 000 ^{14}C years Reconstructed from Coleopteran Remains

Data source: BugsCEP

6.7.1 Aims

The aim was to reconstruct a general picture of the thermal regime of Europe for the last 20 000 years, in 1 000 year time slices, using the data available from ^{14}C dated samples in BugsCEP. By grouping together samples from a wide geographical area, into broad time slices, it may be possible to simulate the type of assemblages that would be expected where samples from a single site encompass environments or climates that change faster than the resolution of the samples. It should then be possible to evaluate the ability of the jackknife method to identify such assemblages, and thus provide indications of the reliability of a particular temperature reconstruction. The reconstruction is compared with existing published reconstructions, and the reliability of the former discussed, with respect to thermal gradients and the representativeness of the time slice assemblages. A habitat based

reconstruction is presented for the same data, in the interest of identifying general trend in the evolution of European environments.

6.7.2 Introduction

A number of biological proxy derived reconstructions of the past 10-20 thousand years of European climate have been undertaken (e.g. Davis *et al.*, 2003) at a variety of resolutions, and overviews can be found in a number of general texts (e.g. Battarbee, 2004; Bell & Walker, 2005). Similar reconstructions from Coleoptera have been undertaken by Coope & Lemdahl (1995), Coope *et al.* (1998) and others. The aim here is to see how a general European millennial scale climate reconstruction using the sites in BugsCEP and the BugsMCR utility is similar to earlier reconstructions. BugStats is also used to produce a general environmental reconstruction for the same period and area, and the implications of this discussed in relation to the variability in the thermal reconstructions. In addition, a correlation coefficient is used to assess the similarity of faunas between time slices. For a more detailed description of the mechanics of the reconstructions see chapters 4 (BugStats) and 5 (BugsMCR). All dates are given in ^{14}C years.

A naive approach to the dating of sites is used here, with no age depth curves being constructed and only samples with direct ^{14}C dates being used. No attempt is made to assess the differential reliability of the dates, and it is possible that a number of bulk dates are included. The broad nature of the time slices may help to minimise the effects of these. The implications of individual sites for reconstruction will not be discussed in detail, and this section should be considered more of an experiment than an absolute description of the past 20 000 years of climate and environmental change.

6.7.3 Sites

Queries were created within BugsCEP to limit the sites extracted to those within European countries, with Greenland excluded due to its distance from mainland Europe and the potential for the Greenland icecap and Atlantic Ocean making any temperature reconstructions significantly different from those of mainland Europe. The position of the Fennoscandian ice sheet will also have had a significant impact on the climate of various time slices (Coope *et al.*, 1998), and due to the geographical range of sites used this impact will vary. Figure 6.26 shows the geographical distribution of 77 sites included in the reconstruction, and Table 6.15 the site names, and time slices that they represent. This is coincidentally the same number of sites as used by Coope *et al.* (1998) to investigate temperature gradients in northern Europe, although the list of sites is not identical, with only 28 sites being common to the two reconstructions. A further 23 sites are present in BugsCEP but either their species lists or dating evidence have yet to be entered fully, and the model will be updated in the near future^{xvii}. The use of age-depth curves for partially dated sites could have increased the size of the dataset considerably, but would have required closer examination of the individual sites and lies beyond the scope of this thesis. The author is reluctant to apply a generic linear age-depth curve to all sites due to the potentially variable nature of the sediments sampled, and the lack of sample depths for a number of important sites in BugsCEP. Such an approach was undertaken by Davis *et al.* (2003) using European pollen data, with a dataset which allowed for a considerable amount of spatio-temporal interpolation. It should be noted that the dominance of sites in the British Isles will undoubtedly have influenced the reconstructions, as will the few more southerly sites (Figure 6.26 and Figure 6.27). This is, of course, not an ideal distribution of sites for a single reconstruction, but they are used here to illustrate a number of points for discussion. Although Coope *et al.* (1998) demonstrated both latitudinal and longitudinal gradients in glacial-Holocene transition temperatures, only a plot of the latitude of sites per time slice is provided here (Figure 6.27). Factors associated with latitude are controlling factors in biodiversity (Mittelbach *et al.*, 2007), and the latitudinal spread of sites may therefore affect the environmental reconstructions. It is important to remember that, due to the nature of the method, the MCR temperature reconstructions will represent the mutual range of the

^{xvii} 13 of the sites used by Coope *et al.* (1998) use unpublished data that has not yet been published or acquired for inclusion in BugsCEP.

greater part of the fauna from all latitudes in a time slice, and not the average temperature represented by them. Altitude is not considered in the analyses, as the data was not available from all sites, and may play an important role in the differences in reconstructed temperatures. It is hoped, however, that the large time slices, and the lack of reliance on 100 % thermal overlaps will smooth out any regional differences.

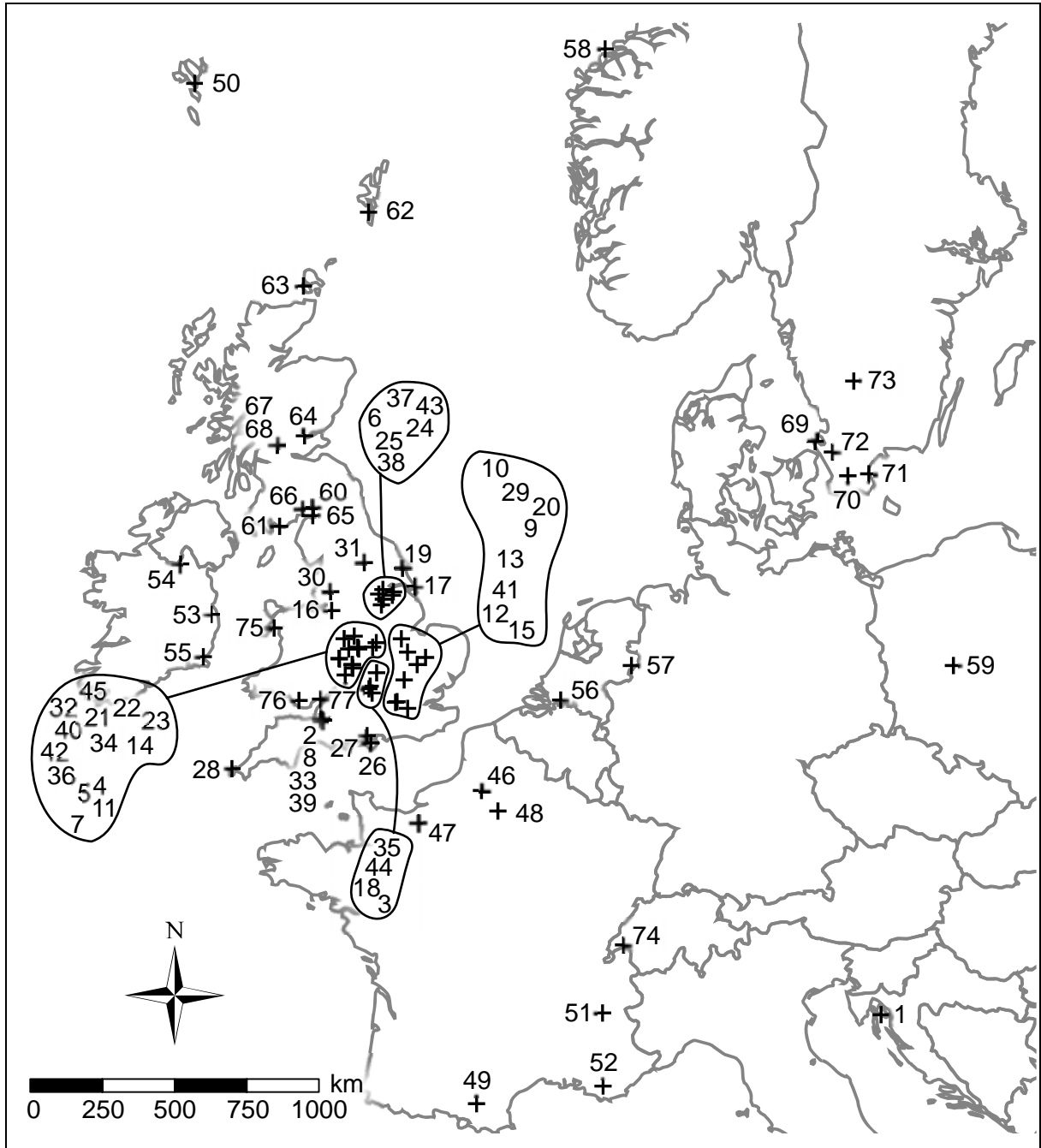


Figure 6.26. Map of sites used in the 20 000 ¹⁴C year reconstruction, see Table 6.15 for site names and time slices. The positions of the numbers in the call-out bubbles show the approximate position of the respective site symbols (+) on the map.

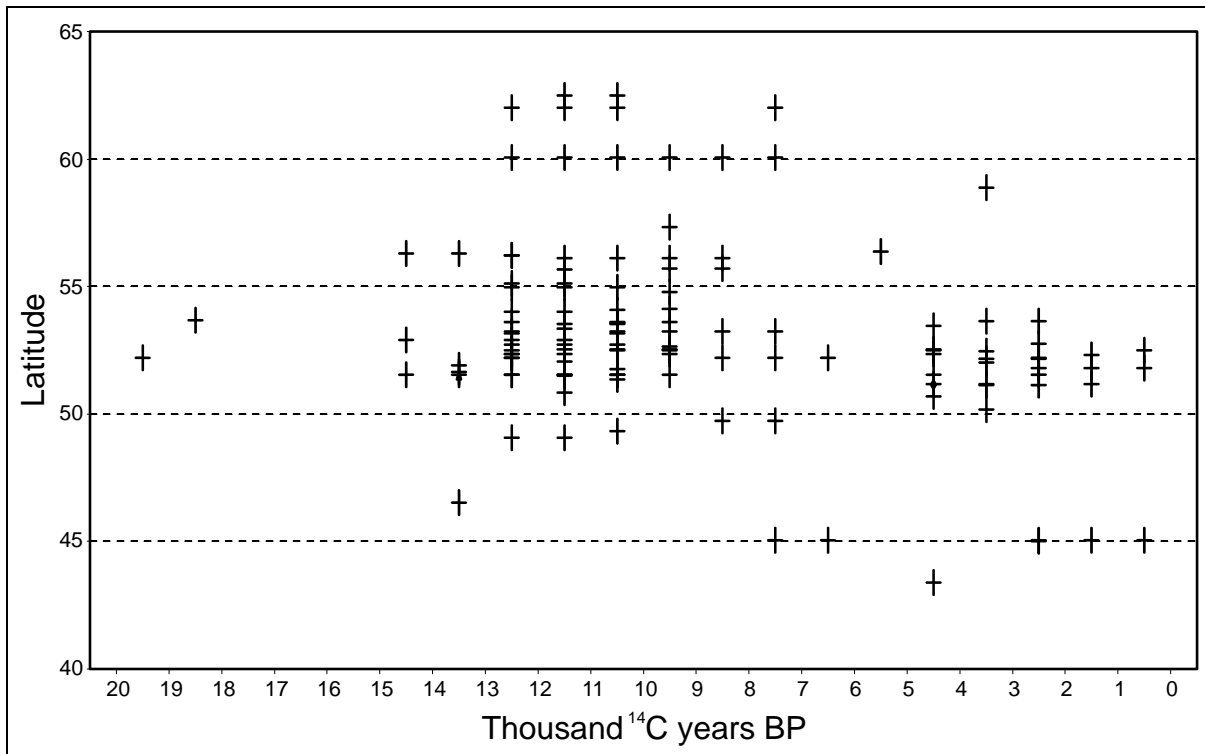


Figure 6.27. Plot of latitudes of sites in the 20 000 ¹⁴C year reconstruction, per time slice.

Table 6.15. List of sites used in the 20 000 ¹⁴C year, 1 000 year time slice reconstruction, indicating the time slices for which each site has ¹⁴C dated samples. Note the absence of time slices 18-17K, 17-16K and 16-15K which produced no data. See BugsCEP for references for all sites.

No.	Site Name	Country	19-20K	18-19K	14-15K	13-14K	12-13K	11-12K	10-11K	09-10K	08-09K	07-08K	06-07K	05-06K	04-05K	03-04K	02-03K	01-02K	00-01K
1	Okruglo	Croatia															X		
2	Abbots Way	England												X	X			X	
3	Abingdon	England				X													
4	Alcester: Alluvial	England									X	X	X						
5	Alcester: Coulter's Garage	England															X		
6	Armthorpe	England						X											
7	Aston Mill: Late Holocene	England														X			
8	Baker site	England												X					
9	Barnwell Station	England	X																
10	Baston Fen	England					X	X											
11	Bidford on Avon: Pilgrim Lock	England													X	X			
12	Colnbrook	England				X													
13	Colney Heath	England				X													
14	Croft	England							X						X				
15	Croydon	England							X										
16	Davenham: Church Moss	England					X	X	X	X	X								
17	Dimlington	England		X															
18	Farmoor: Late Glacial	England						X											
19	Gransmoor	England					X	X											
20	Isleham	England												X					
21	Lea Marston A	England						X											
22	Lea Marston B	England								X									
23	Leicester: Early Holocene	England								X									
24	Messingham	England							X										
25	Misterton Carr	England												X					

No.	Site Name	Country	19-20K	18-19K	14-15K	13-14K	12-13K	11-12K	10-11K	09-10K	08-09K	07-08K	06-07K	05-06K	04-05K	03-04K	02-03K	01-02K	00-01K
26	New Shide Bridge	England												X					
27	Northmoor LG	England						X											
28	Porth Meare Cove	England														X			
29	Ramsey Heights	England														X			
30	Red Moss	England					X		X	X									
31	Ripon	England								X									
32	Rodbaston Hall	England							X										
33	Rowlands Track	England													X				
34	Shustoke	England													X				X
35	Sproughton	England						X											
36	Stourport	England																X	
37	Thorne Moors Trackway Site	England														X	X		
38	Tiln	England						X											
39	Tinney's Brushwood	England														X	X		
40	West Bromwich	England					X		X	X									
41	West Drayton	England						X											
42	Wilden	England					X	X		X									
43	Yarborough Quarry	England							X										
44	Yarnton	England															X	X	X
45	Yoxall Bridge	England															X		
46	Conty	France					X	X	X										
47	Havre	France									X	X							
48	Houdancourt	France					X	X											
49	La Borde	France							X										
50	Skalafjordur	Faroes										X							
51	Taillefer Massif	France									X	X					X	X	X
52	Tourves	France													X				
53	Ballybetagh	Ireland					X		X										
54	Drumurcher	Ireland							X										
55	Shortalstown	Ireland					X												
56	Notsel, Mark Valley	Netherlands					X	X	X										
57	Usselo	Netherlands					X												
58	Godøy	Norway						X	X										
59	Zabinko	Poland					X												
60	Bigholm Burn	Scotland						X											
61	Brighthouse Bay	Scotland								X									
62	Clettnadal	Scotland					X	X	X	X	X	X							
63	Heldalewater, Hoy	Scotland														X			
64	Kinfauns	Scotland												X					
65	Redkirk Point	Scotland					X	X	X										
66	Roberthill	Scotland					X												
67	Teith Valley	Scotland					X												
68	Torrie Borehole	Scotland					X												
69	Björkeröds Mosse	Sweden			X	X													
70	Bysjön	Sweden						X											
71	Hanobukten	Sweden								X	X								
72	Körslättamossen	Sweden						X	X	X	X								
73	Ranstad	Sweden								X									
74	Lausanne	Switzerland				X													
75	Glanllynau	Wales			X		X	X											
76	Llanilid	Wales			X	X	X	X	X	X									
77	Redwick	Wales													X		X		

6.7.4 Methods

The BugsCEP database was used to extract abundance data for the past 20 000 ^{14}C years, in 1 000 year time slices, using only ^{14}C dated samples from stratigraphic sequences. Only data from European sites, with the exclusion of Greenland, were used. A total of 3 768 taxon occurrences were retrieved, populating the time slices as shown in Table 6.16. The effective meaning of the cell values are explained in Table 6.17.

Although some samples have more than one ^{14}C date, only one instance of the affected taxa occurrence is included in the calculation lists, thus preventing over representation of the taxa concerned. Due to the presence/absence nature of the method, MCR reconstructions are unaffected by this, whereas the abundance weighted environmental reconstructions are extremely sensitive, and would give grossly exaggerated habitat signals for samples with more than one date.

Table 6.16. Summary of samples and species occurrences from 1 000 year time slices for the past 20 000 ^{14}C years. An ‘occurrence’ is a fossil record of a specific taxon in a specific sample, and may either be an abundance or presence value, see Table 6.16 for more details. Note the important difference between the number of taxa (NSpec), and the number of taxa available for MCR (No. MCR Taxa).

Time slice	No. Sites	No. Samples	Occurrences	No. Taxa (NSpec)	No. MCR Taxa	Total Abundance	MCR Abundance
00-01K	3	3	265	241	50	772	146
01-02K	4	6	376	258	65	1288	230
02-03K	9	11	923	480	95	3328	623
03-04K	8	11	527	258	54	2044	429
04-05K	10	10	435	274	62	1186	191
05-06K	1	1	4	4		4	
06-07K	2	2	66	65	10	226	35
07-08K	6	8	132	108	26	303	76
08-09K	6	10	152	107	18	264	48
09-10K	13	28	1156	385	95	5154	1327
10-11K	19	26	1302	544	136	7248	2521
11-12K	20	46	2142	428	140	6120	2478
12-13K	20	31	1387	411	140	2979	1033
13-14K	6	8	206	135	43	491	131
14-15K	3	4	48	40	15	63	25
15-16K							
16-17K							
17-18K							
18-19K	1	1	26	13	6	26	12
19-20K	1	5	27	21	9	31	14
Instance Totals			9166	3768	964	31541	9319
Unique Totals	77	186		1306	271		

Table 6.17. Explanation of cell and total values in Table 6.16.

	Time slice	Instance Totals	Unique Totals
No. Sites	Number of sites representing time slice		Number of sites in dataset
No. Samples	Number of samples representing time slice		Number of dated samples in dataset
Occurrences	Number of fossil records in time slice	Total number of fossil records	
No. Taxa (NSpec)	Number of unique taxa per time slice	Sum of time slice numbers	Number of unique taxa
No. MCR Taxa	Number of unique MCR taxa per time slice	Sum of time slice totals	Number of unique MCR taxa
Total Abundance	Total number of individuals in time slice	Total number of individuals in dataset	
MCR Abundance	Total number of MCR individuals for time slice (used in climate reconstruction)	Total number of individuals used in MCR reconstruction	

A climate reconstruction was produced for the times slices using BugsMCR (Figure 6.28). The slices were also subjected to a jackknifed MCR reconstruction in order to see the effect of removing one species at a time on the reconstructed temperatures. This could provide information on the internal reliability of each MCR reconstruction by highlighting the degree of variation in the samples. Additional counts and statistics, which are not standard BugsMCR outputs, are provided with each temperature diagram to support the discussions.

The modified Sørensen correlation coefficient (Southwood, 1978) was calculated to provide an indication of the faunal similarity between slices (Table 6.18). This was run using all taxa, and the method is abundance sensitive, which may have the effect of enhancing the difference between low and high abundance slices. The comparison with independent faunal groupings earlier in this chapter (section 6.2.5) suggests, however, that the resultant groupings are viable when compared to those delimited on the basis of ecological knowledge.

A Bugs EcoCode based environmental reconstruction was performed using a variety of settings. Only standardized, as a percentage of the sum of representations (%SumRep), diagrams are presented, as the high variation in species numbers and abundance values between samples rendered the other diagrams practically useless for the comparison of habitat indications between time slices. Calculations were performed using both ‘all taxa’ and ‘species level identifications only’ settings, in order to identify variations in the effect of higher level identifications on the reconstructions. It is possible that generic level identifications are proportionally more common in samples from colder or more disturbed periods, and that their removal will reduce the environmental resolution of these samples. See Chapter 4 for a full explanation of the reconstruction settings and their implications.

6.7.5 Results and discussion

As explained in Chapter 5, BugsMCR does not employ the calibration/correction method of Atkinson *et al.* (1986), and outputs a range of temperatures in which all values are equally probable on the basis of the assemblage underpinning the reconstruction. These ranges present a more realistic reconstruction of large time slices than a corrected single value, due to the fact that temperatures will have undoubtedly varied during the time slice, and that the single value cannot express this. As yet, the correction method is not able to provide sample based error estimates, and so standard deviations are not available to simulate the range of values derived from a raw MCR calculation. Despite this, the corrected values of previous authors (e.g. Coope *et al.*, 1998) should fall within the range of uncorrected MCR values, and thus be comparable to an extent. A number of other authors have presented both corrected and uncorrected data, which makes comparison easier. Although there are a

number of interesting differences between the reconstruction presented here and a number of previously published temperature curves, the majority of time slices coincide with the ranges of previous publications. The regional comparison presented by Coope & Lemdahl (1995), which illustrated local variations in Lateglacial-early Holocene climate change, may help in explaining some of the differences between the general reconstruction and specific regional curves.

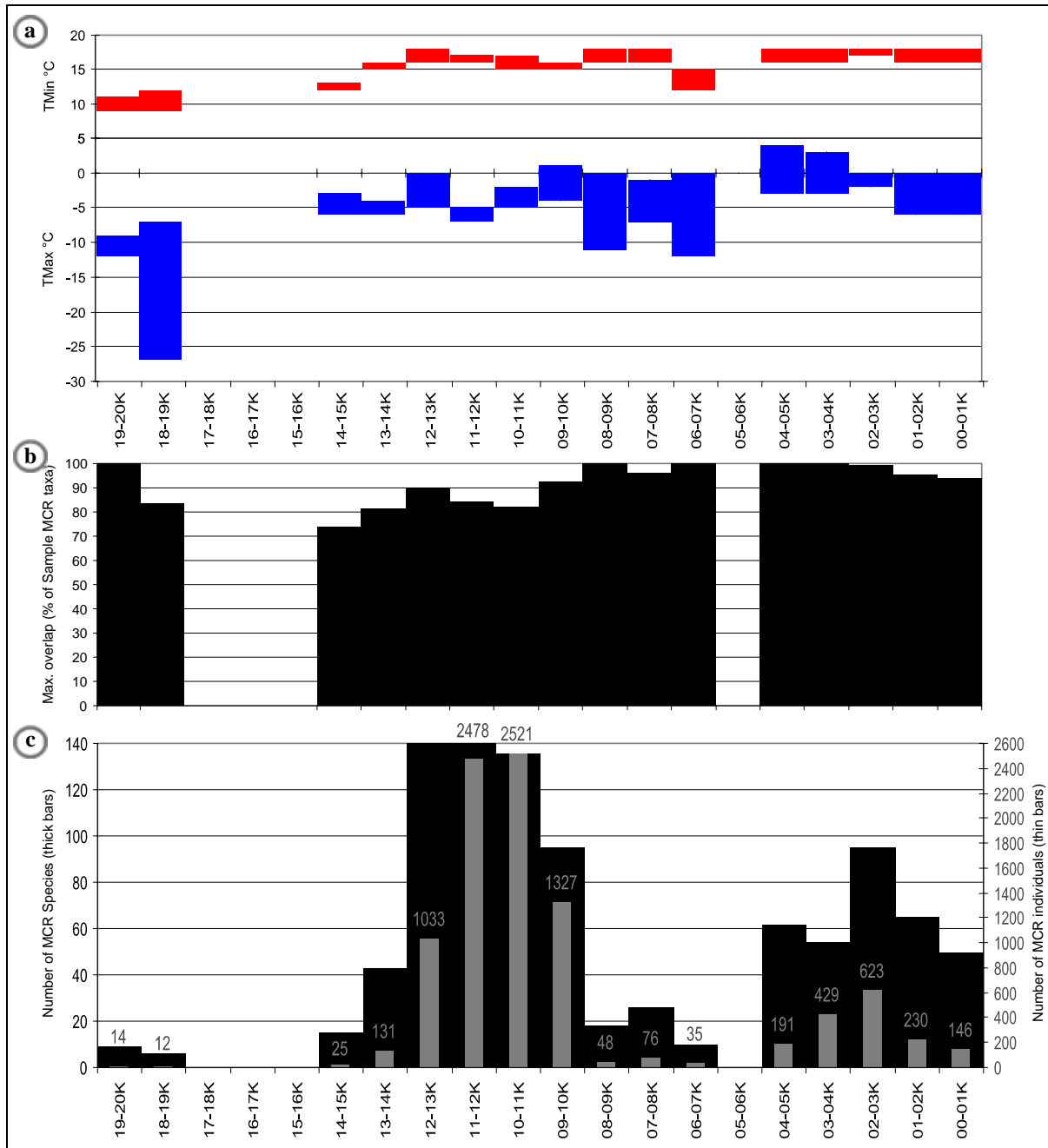


Figure 6.28. MCR temperature reconstruction for the last 20 000 ¹⁴C in 1 000 year slices using all European ¹⁴C dated stratigraphic sequence samples in BugsCEP (Greenland excluded). (a) Bars represent the possible range of temperatures as indicated by the beetle faunas, for TMax and TMin; (b) the percentage of species in the area of maximum overlap from which the MCR values are calculated; (c) number of taxa (thick black bars), and individuals (thin grey bars) used in the reconstruction.

Despite the large number of species involved (up to 140 in time slices 12-13K and 11-12K), at least 75 % of species represented the area of greatest overlap in each of the time slices that included MCR taxa (Figure 6.28b). This suggests that, despite the large geographical spread of sites, the majority of

the faunas represented are relatively thermally homogenous. Although the time slices are large, and arbitrary in terms of their boundaries, a degree of comparability is displayed between Figure 6.28a and earlier reconstructions, although it differs between TMax and TMin.

Glacial-early Holocene transition

Whilst TMax shows a progressive rise in summer temperature values from 12-13°C to 16-18°C between 15 ka and 13 ka BP, TMin does not. The Lateglacial-early Holocene TMax reconstruction is in reasonable agreement with the British Isle reconstructions of Atkinson *et al.* (1987) and Coope & Lemdahl (1995), although the range of temperatures reconstructed for 13-14K are towards the upper limits of what would be expected. TMin, on the other hand, fails to resolve the expected lower temperatures of the 14-15K time slices. It also indicates a significant dip in winter temperatures between 12 ka - 11 ka BP, followed by a rise to the warmest winter temperatures reconstructed for the Early Holocene (09-10K), a pattern which is not mirrored by TMax. It would appear that the thermal range of the sites used disguises the Younger Dryas as identified in the British fossil beetle record to 11-10 ka BP (Atkinson *et al.*, 1997), or for some reason makes it apparent in the earlier time slice. It may be possible to explain this in terms of the presence of Scandinavian sites in these sample, which, probably due to the close proximity of the Fennoscandian ice sheet, do not synchronously show the same climate development as the British Isles for the period 13 ka to 9 ka BP (Coope & Lemdahl, 1995). The Bølling/Allerød Interstadial, c. 13-11 ka BP (Bell & Walker, 2005), is well reconstructed with the warmer 12-13K and colder 11-12K time slices corresponding to the Bølling and Allerød respectively. Jackknifing (Figure 6.29a) adds little to the interpretation of these slices, the reconstruction being apparently quite consistent through the sites when compared to earlier slices. It could also be that the geographical range of the sites, and thus variations in the chronology and magnitude of climatic shifts is too great to be identified by the removal of any single species. The multiple removals variation of the method (see Chapter 5) may prove more useful in such cases.

Examination of the jackknife data does, however, shed light upon the overall development of European winter temperatures during the period 15 ka-12 ka BP, and enhances the potential of the TMin values to illustrate the trend seen for TMax. Closer examination of the 14-15K time slice may help to explain the results. The individual removal of four of the species from the reconstruction result in significant downward expansion of the lower limit of TMin, although at first glance it is surprising that two of these particular species lead to the possibility of a colder reconstruction. *Boreaphilus hennigianus* Sahl. and *Bembidion fellmanni* (Mann.) are both cold, stenothermic species according to their MCR data, both with TMax tolerances of 6-13°C, and so one would logically assume that their removal from calculations would allow other, less cold tolerant species to draw the reconstructed temperatures upwards, as is the case with the TMax values. The exclusion of the warm, eurythermic species *Hygrotus inaequalis* (F.) and *Chaetarthria seminulum* (Hbst.) results, as would predicted on this premise, in the downward extension of the reconstructed temperatures. Closer examination of the climate space map for the time slice shows a complex set of overlays (Figure 6.30), and suggests that at least two conflicting faunal components are involved, and from Table 6.15 we could suppose that this may be the result of the geographical location of the sites. Of the three sites, Glanllynau and Llanilid in Wales may represent a relatively warmer, more oceanic climate, whereas Björkeröds Mosse, in Sweden, may be responsible for the colder climate strongly influenced by the Fennoscandian icesheet. However, Coope *et al.* (1998) detected no such west-east gradient at this time, and the situation is not so simple. In fact, for the 14-15K time slice, none of the above mentioned species occur at Björkeröds Mosse, all of them occur at Glanllynau, and all with the exception of *B. fellmanni* occur at Llanilid. From examination of the MCR reconstructions for each site it can be suggest that the dating evidence for Llanilid, a humic substances date, may be unreliable, as the sample gives a TMax reconstruction of 17-21°C. Björkeröds Mosse is represented by only two species, *Amara alpina* (Payk.) and *Agabus bipustulatus* (L.), and five individuals in this time slice. Clearly, further well dated sites are needed for a detailed analysis.

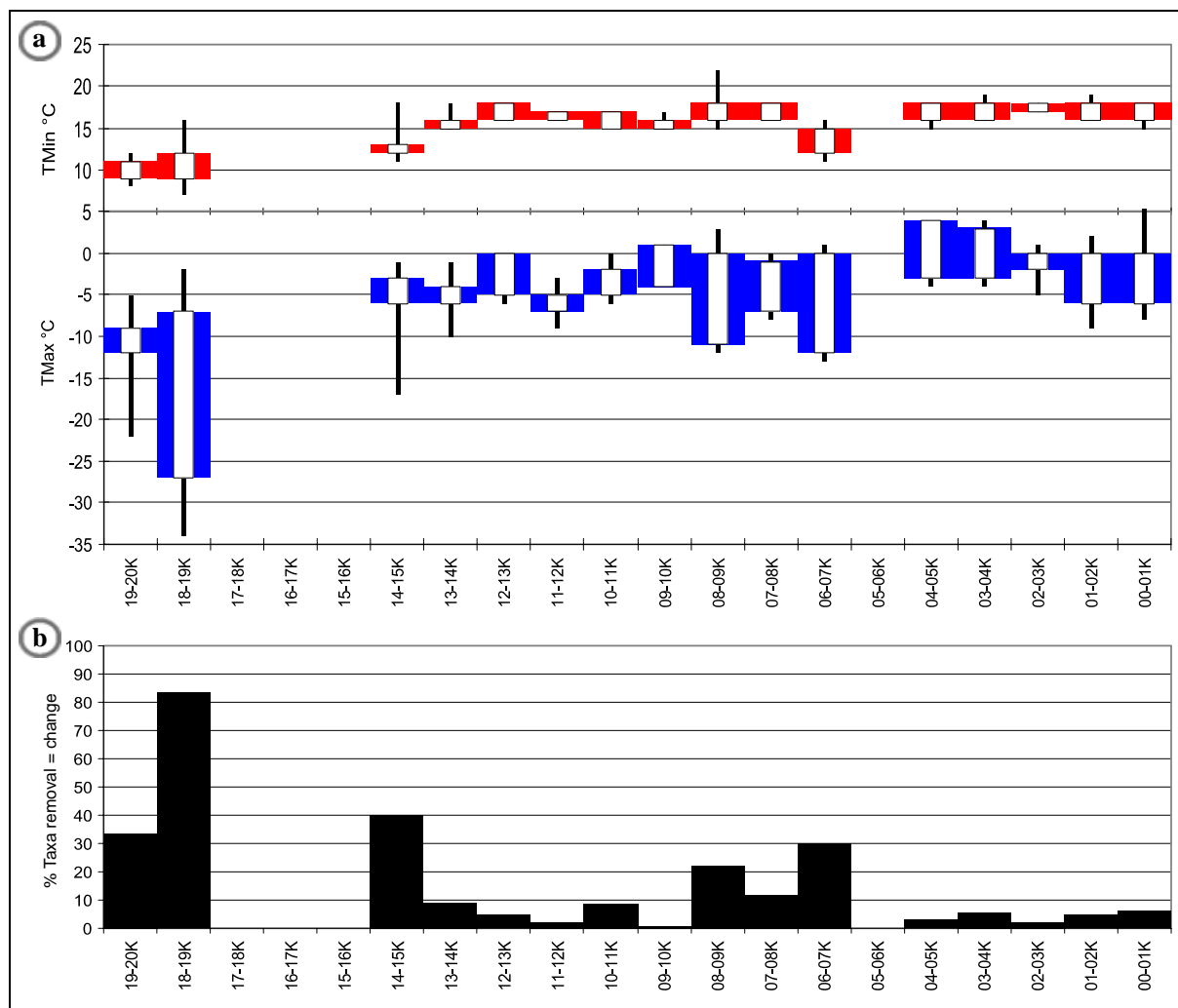


Figure 6.29. Jackknife results from 20ka temperature reconstruction, showing (a) the maximum (black extension bars) enlargement, and reduction (white boxes) of combined envelopes caused by the removal of any one taxa. The standard MCR values are shown as in Figure 6.28a. (b) shows the percentage of taxa whose removal causes a change in the reconstructed temperature range.

Both TMax and TMin realistically identify the significantly colder 20 ka – 18 ka period, with results comparable to other reconstructions of 9-12°C summer temperatures. The two millennia are represented by single sites in the British Isles, and include assemblages which represent temperatures of between 5-9°C lower than the reconstructed summer temperature for the last 1 000 years. Ideally, more data should be included in these time slices to provide a more robust reconstruction.

Fixed 1 000 year slices, rather than slices clustered by thermal characteristics, or individual sample by sample reconstructions (e.g. Atkinson *et al.*, 1987; Coope & Lemdahl, 1995), increase the probability that periods of rapid climate change will be missed in the reconstructions. Even though the 12-13K to 11-12K, and 10-11K slices could, on the basis of our current understanding of Lateglacial-early Holocene climate change, have coincided with the Bølling/Allerød and Younger Dryas periods respectively, little evidence was found for them in this reconstruction other than a slight indication in TMax values (Figure 6.28a). The changes in the percentage of species representing the area of maximum overlap (Figure 6.28b) may provide a proxy for the degree of climate change in a time slice, in that slices that encompass greater change are more likely to contain assemblages representing more thermally diverse faunas. Another possibility is that some of the variation is caused by changes in geographical thermal gradients over time, as described by Coope *et al.* (1998), or lags in climate change across Europe.

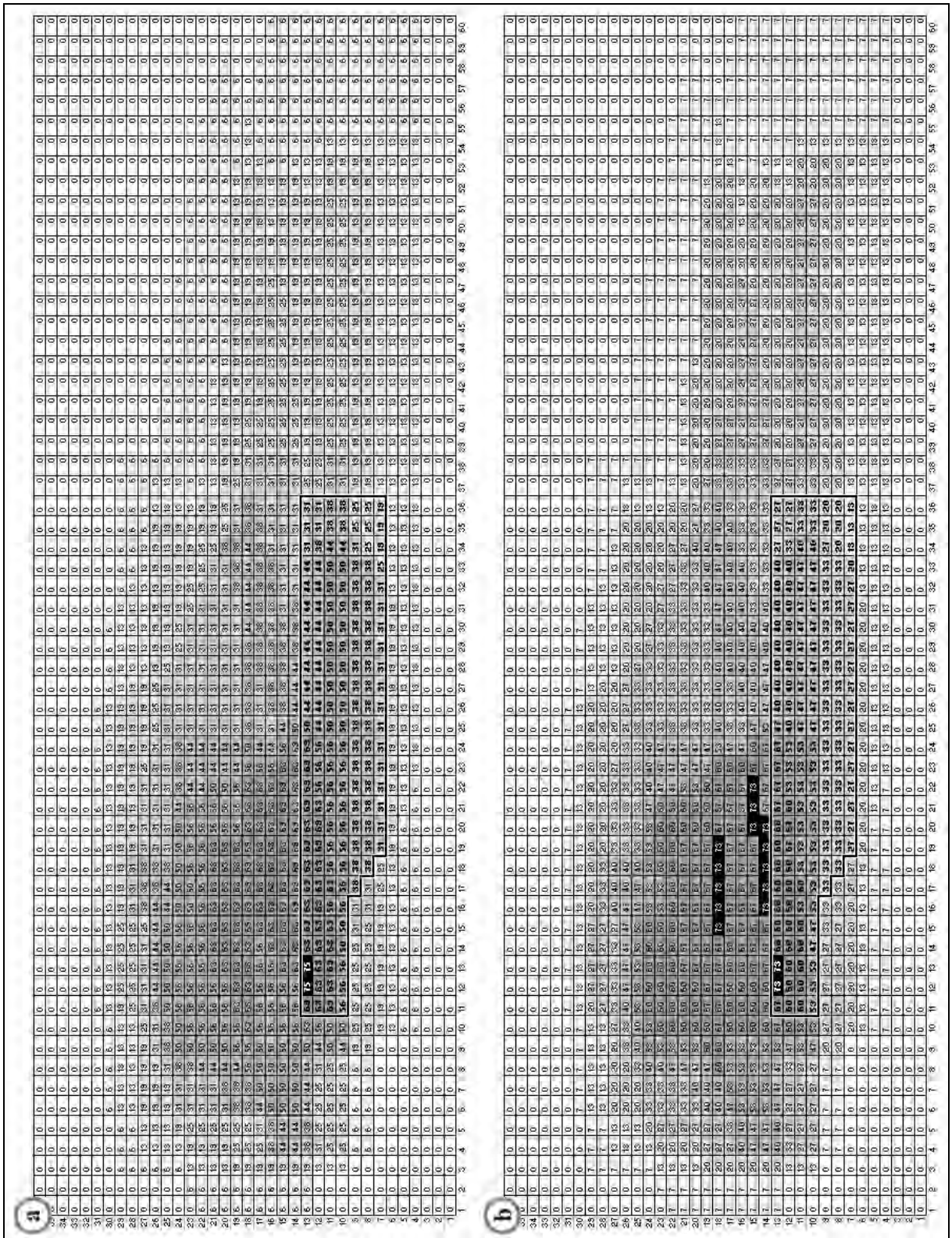


Figure 6.30. Climate space maps for time slice 14-15K, showing percentage overlaps for (a) all taxa, and (b) with *Boreaphilus henningianus* Sahl. excluded, the thermal envelope of which is outlined and shown in bold. The cells with maximum overlap, which make up the mutual climatic range are shown with white text on black cells. Note how the removal of *B. henningianus* creates a discontinuous area of maximum overlap, the outer limits of which give the MCR temperature values. This illustrates a potentially split fauna.

Early-mid-Holocene

The time slices from 8 ka to 5 ka contain considerably less data than their neighbouring time slices (Figure 6.28c), which results in less constrained TMin values, suggesting a probability of colder winters. This seems unlikely on the basis of earlier reconstructions, and is most likely an artefact of the small dataset. Comparison with time slices 19-20K and 18-19K, which are based on even less data, illustrate the ability of MCR to give acceptable temperature reconstructions with even small numbers of species, although the lack of error estimation is a problem. The dating evidence for the single site in time slice 6-7K, 'Alcester: Alluvial' (Shotton *et al.*, 1977), is unreliable and the sample should most probably be moved into the 8-9K time slice. The lack of dated samples in the Early to mid-Holocene is clearly a problem. Interpolation of dates, by the application of age-depth curves would help increase the size of the dataset, but introduces another set of errors which could make the results less reliable. In this respect, BugsCEP can be used to identify areas in need of further research, and this is clearly one of them.

The jackknife results (Figure 6.29a) suggest that the reconstruction of the 8-9K time slice could be made warmer by the removal of *Patrobis assimilis* Chaud., a eurytopic carabid whose thermal envelope suggests a slightly eurythermal, but cold tolerance. The majority of other species in the time slice are either warm stenotherms or eurytherms, and suggest a wider extension of the MCR into warmer climates. *P. assimilis* is found at two of the six sites representing this time slice, Clettnadal on Shetland and Hanobukten in Sweden, but only represented by one individual at each site. Obviously its presence cannot simply be dismissed on the grounds that it does not fit the pattern of the other species, just as a radiocarbon date must be explained rather than dismissed. It is possible that the species is present at one extreme of its tolerance range, or that it is a casual occurrence at both sites. In either case, its effect on the temperature reconstruction cannot easily be dismissed without treating species with equally low abundances in other time slices in the same way. The implications of filtering out low abundance species are not simple, and there is a risk of losing information, especially from low frequency indicator species.

Mid-late Holocene

The MCR reconstruction from 5 ka BP to present appears to present a consistent picture of a stable late Holocene climate. The data do suggest a swing in the variability of winter temperatures, from warmer between 5 ka and 3 ka, and colder from 2 ka to the present day, although it should be remembered that there are no modern samples in the dataset. The 00-01K time slice is poorly represented, by only three sites, and may not present a particularly reliable picture of this segment when compared to the more data rich ones. The addition of modern reference sites, and even a higher resolution beetle based MCR analysis of the past millennium, would be a worthwhile exercise.

Relationships within the data

The compilation of large datasets provides opportunities to investigate commonly cited statistical relationships in populations, and their applicability to fossil assemblages, as well as relationships between the output data from BugsMCR. It was stated earlier (Chapter 5) that the percentage of species whose removal leads to a change in temperature values, as a result of the jackknife method, could be a valuable indicator of reliability in reconstructions, possibly in the form of a measure of the amount of variability in an assemblage. It is commonly accepted that the number of species used in a reconstruction is a controlling factor on the reliability of any reconstruction – the more species, the greater the evidence. An approximate log-log relationship is seen between these two variables in the 20 ka dataset Figure 6.31a, suggesting that the 'percent of species removals leading to changed temperatures' variable (PctSppTDiff) may not contribute anything beyond the *number of species* variable (NSpec). The data points become more dispersed towards the higher number of species, and the relationship needs investigating more thoroughly with site datasets, rather than using compiled pseudo-samples such as the above time slices.

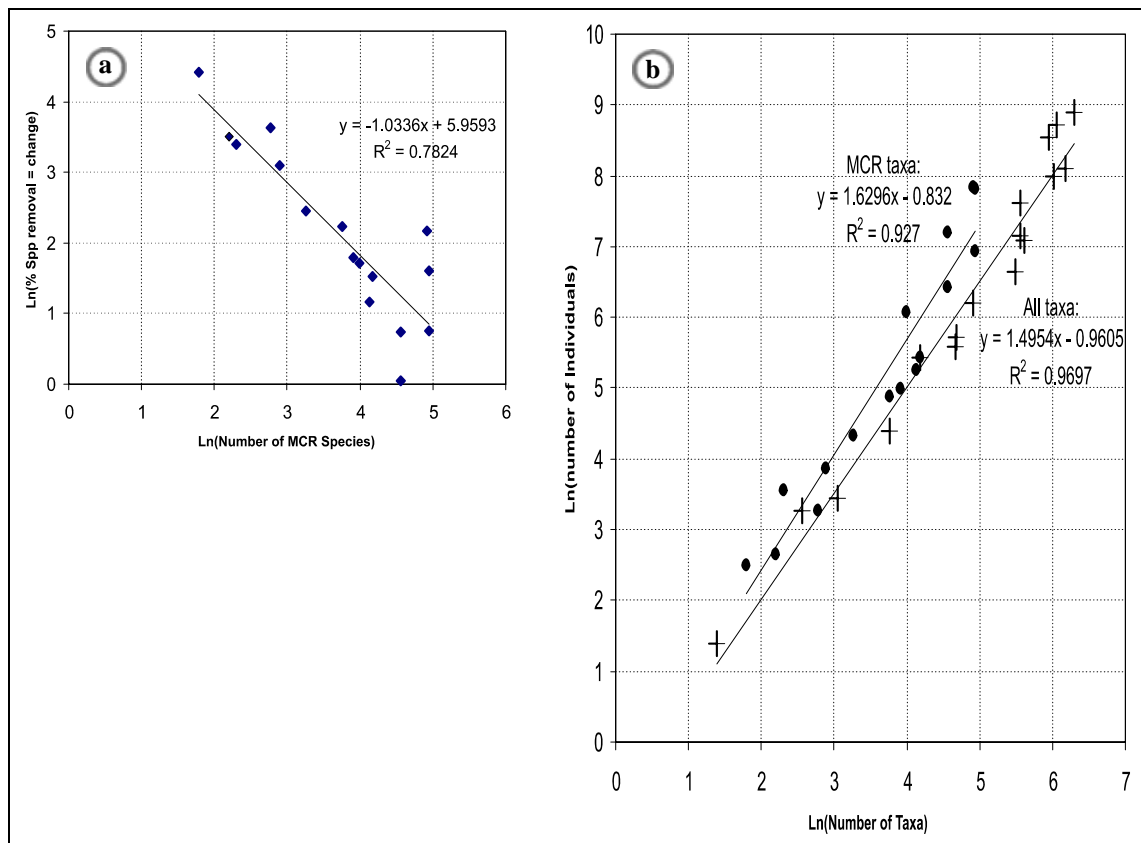


Figure 6.31. Graphs showing relationships, in the 20 ka dataset, between (a) the number of MCR species and the percentage of species whose removal leads to a change in temperature values; and (b) the number of taxa and number of individuals for the MCR subset and full dataset.

The relationship between the number of species and the number of individuals in animal populations (the S:N ratio) has been the subject of attention for many years, and a number of descriptive models exist (Southwood & Henderson, 2000). The implications of these relationships for environmental reconstruction from fossils were briefly discussed in Chapter 4. In the 20ka dataset, both the MCR species and full dataset exhibit ln-ln relationships between number of taxa and individuals (Figure 6.31b). Due to the nature of the data selection, this is in fact the relationship between the sum of (^{14}C dated) samples in each time slice, and may not represent a true S:N ratio for the individual samples. The fact that the relationship for the MCR taxa is extremely similar to that for all taxa, suggests that the subsetting of a species list to only MCR taxa does not introduce bias by changing the S:N relationship. The BugsCEP database provides an excellent opportunity to study these relationships for fossil assemblages, and more work should be done on this.

Time slice correlations

From Table 6.18 it can be seen that there are two, or possibly three clusters of similar time slices according to the modified Sørensen coefficient (see Chapter 4 and Southwood, 1978). Slices 2-3K and 3-4K are most similar, which suggests that the period 4 ka to 2 ka BP was faunally, and thus environmentally, relatively stable. Only two of the 14 sites representing this time period are outside of the British Isles, however, so the pattern may only be valid for this small area of Europe. It is very easy to draw far reaching conclusions from correlation data, but immense care should be taken in assigning cause and effect, or even significance to the correlations. The relative faunal similarity of the past two millennia, preceded by a reduction in similarity with the third millennium BP, could be considered to reflect recent human impact on the environment. The Late Bronze Age-Early Iron Age is often considered the starting point for large scale landscape modification, especially in the British Isles (cf. Bell & Walker 2005), where the majority of the beetle data used here are from. However tempting

this may be, the number of samples involved are too few, and a more detailed consideration of chronologies and site locations needs to be employed before this can be further investigated.

Table 6.18. Modified Sørensen (Southwood, 1978) correlation coefficients comparing 1 000 year time slices for the past 20 000 years. Note that the full dataset has been used, and not just the MCR species, and time slices with no data have been omitted.

	00-01K	01-02K	02-03K	03-04K	04-05K	05-06K	06-07K	07-08K	08-09K	09-10K	10-11K	11-12K	12-13K	13-14K	14-15K	18-19K
00-01K																
01-02K	0.49															
02-03K	0.25	0.35														
03-04K	0.11	0.17	0.73													
04-05K	0.27	0.25	0.28	0.27												
05-06K	0.00	0.00	0.00	0.00	0.00											
06-07K	0.14	0.15	0.09	0.06	0.07	0.00										
07-08K	0.15	0.15	0.09	0.07	0.13	0.00	0.44									
08-09K	0.13	0.12	0.08	0.07	0.18	0.00	0.22	0.47								
09-10K	0.10	0.16	0.29	0.22	0.18	0.00	0.03	0.06	0.06							
10-11K	0.09	0.12	0.31	0.23	0.15	0.00	0.03	0.04	0.04	0.39						
11-12K	0.10	0.14	0.31	0.25	0.14	0.00	0.03	0.05	0.05	0.32	0.59					
12-13K	0.15	0.20	0.33	0.27	0.23	0.00	0.06	0.09	0.09	0.37	0.38	0.47				
13-14K	0.11	0.09	0.07	0.06	0.09	0.00	0.08	0.12	0.17	0.05	0.08	0.08	0.12			
14-15K	0.09	0.06	0.03	0.03	0.06	0.00	0.07	0.06	0.10	0.02	0.02	0.02	0.04	0.19		
18-19K	0.01	0.01	0.00	0.01	0.01	0.00	0.06	0.05	0.06	0.00	0.01	0.01	0.01	0.07	0.15	
19-20K	0.01	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.00	0.01	0.01	0.01	0.06	0.09	0.21

The similarity between time slices 11-12K and 10-11K is unexpected, as these should, in theory, represent the climatically different Bølling/Allerød (warm) and Younger Dryas (cold) periods respectively. Previous work (e.g. Coope *et al.*, 1998) has shown that, at least in the British Isles, beetles can be used to identify these climatic zones, and as discussed above, it is surprising that they are not easily identified in this study. Even so, the time slices described quite different thermal environments. Of the 695 taxa found in these slices, 277 are common to both, but only 89 of these are in the MCR dataset. It appears that other faunal components than those in the MCR calibration set are more stable than the beetles used to examine climate change. The relatively high coefficient value indicates that climatic difference is no guarantee of a difference in the correlation coefficient. This difference is to be anticipated, as the MCR method and coefficient calculations are mathematically dissimilar, including the fact that the coefficient uses abundance data whereas MCR does not. The coefficient of comparison, when based on raw assemblage data, seems to be of little use in defining climatic zones, even if it may be useful when supporting zones based on the ecology of the beetles (see section 6.2.4). As yet, too little statistical work has been done on fossil beetle faunas to be able to assess the effect of different types and sizes of assemblages on the ability of the coefficient to help in analyses, and more work is necessary on a wide variety of sites and modern data.

There is too little data to be able to investigate the mid-Holocene from this dataset, and the apparently very low similarity between faunas from 6 ka to 3 ka BP is simply a reflection of this. One might expect stability during this period at the millennial resolution, even if large herbivores are active in keeping forest structures periodically open (Vera, 2000).

There is a weak positive relationship (linear regression: $R^2=0.22$) between the value of the coefficient and the number of taxa (in the older of the two time slices compared in each cell). This could be a reflection of bias in either the coefficient or the dataset, and individual sample based analyses should be undertaken to investigate this, along with the application of other coefficients. Faunal stability is an important component in our understanding of past environmental change, and its study could help refine our understanding of how species react to varying degrees of natural, or human induced, habitat loss.

6.7.6 An experiment in large scale environmental reconstruction, and a test of the BugStats %SumRep standardization method

The full time slice dataset, not restricted to MCR taxa, was used to produce an environmental reconstruction using BugStats. This is not an unproblematic undertaking, and it is used here as an experiment to see how the BugStats system responds to such a problem. The geographical range of sites used is relatively large, from a European perspective (Figure 6.26), and so at any one time there will be a number of different large scale environments at the different sites. Beetles may, in many cases, only represent the local environment of a site, and thus general changes in landscape may not be seen by compiling data from multiple sites. In addition, the large time slices mean that the environments of any site may vary considerably within the time slice.

A number of these concerns are equally valid for the temperature reconstruction, but since the reconstruction of habitats includes assumptions based on considerably more variables than temperature, the results of this experiment are less likely to be accurate. The results are not expected to be viable, but are presented here (Figure 6.32) for the sake of experimentation, and as an aid to the discussion of climate-environment interactions, and the use of classification based habitat reconstructions for large scale environmental reconstruction. If anything, the diagrams may show the evolution of the general environment of northern Europe over the last 20 000 years.

The 05-06K timeslice, with only presence values for four taxa from the site Kinfauns, in Scotland, proved to have far too few species to be of use, even with standardization. These four taxa, *Cerylon* spp., *Grynobius planus* (F.), *Anobium* sp., and *Apoderus coryli* (L.) however, all suggest *Wood & trees* with *Dry dead wood* available, but most probably do not account for 100 % of the habitats, as their inclusion in the diagram would have implied. As indicated above, more dated samples are needed from the Early-mid Holocene to be able to answer a number of key questions on the environment of this period.

The *Stored grain pest* chart has been omitted from the diagram (Figure 6.32), due to the slight nature of the signals given by it. Although there are indications of stored grain pests at various points through the data, the majority of these are high level identifications on genera such as *Ptinus*, which are not specific enough to be of use in this large scale reconstruction. Time slice 02-03K includes a number of individuals of *Sitophilus granarius* (L.) from Okruglo in Croatia (Smith *et al.*, 2006), which represent the earliest dated occurrence of the grain weevil in this dataset. The count data for this site is semi-quantitative, which is not ideal for habitat reconstruction, but as abundance weighting is not being used in this particular reconstruction the data are of some use.

There appears to be a peak in woodland indications during the mid Holocene (6 ka to 3 ka BP), with specific indications of deciduous or coniferous woodland at various points. This is in line with existing understanding of the British landscape of this time, from which the majority of samples originate. The lack of data for the period 9 ka to 5 ka BP may obscure the true nature of this peak. There appears to be a greater presence of aquatic habitats in the early Holocene than mid-late Holocene, and the earliest post-Glacial samples (14-15K and 13-14K) suggest a dominance of running water over standing water. This could be interpreted in terms of the release of water from the melting of the remains of the ice masses and permafrost covering the British Isles, which then stabilises in the formation of lakes. These lakes are successively silted up, and give way to the development of more fully vegetated landscapes with time. This hypothesis can be at least partly supported by the similarity of the *Disturbed/arable* signals which should correspond to more mobile, disturbed environments.

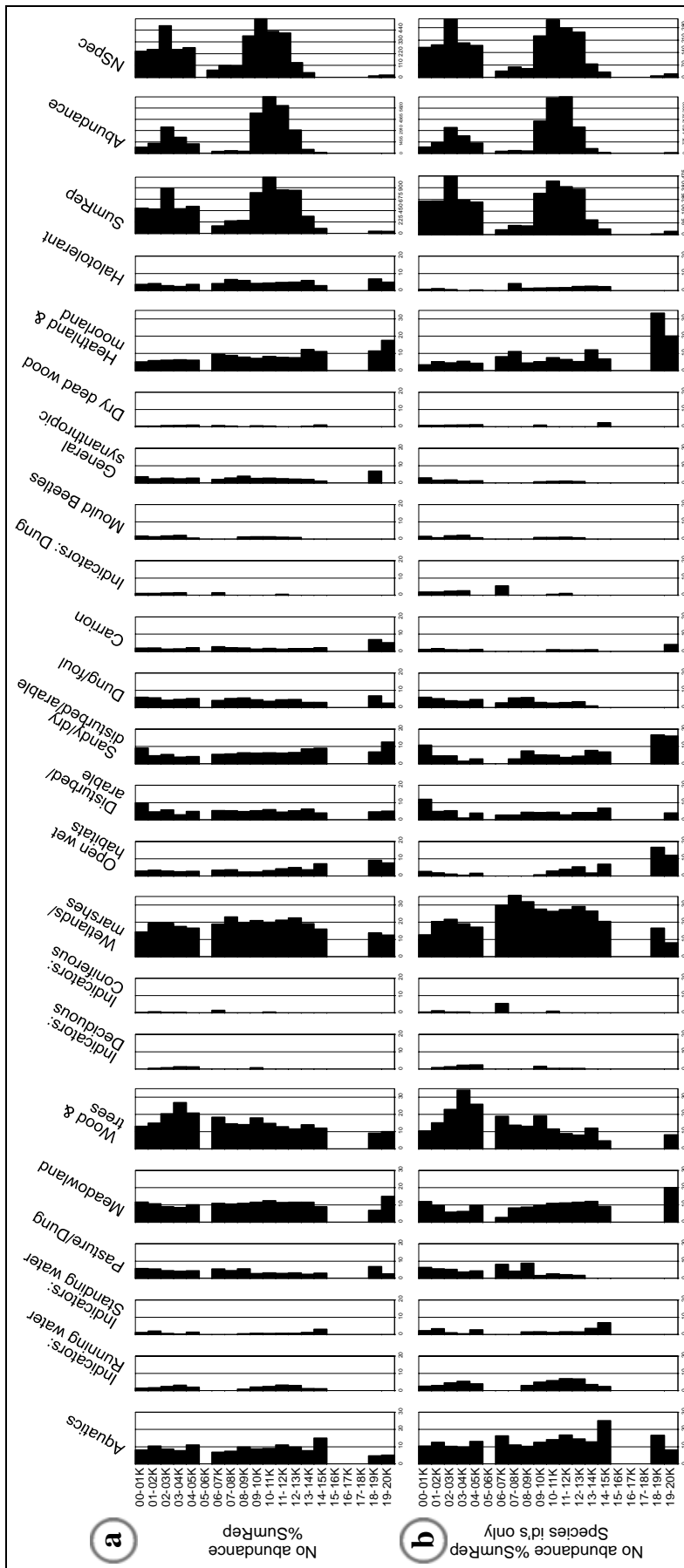


Figure 6.32. BugStats EcoFigs for the last 20 ka in 1 000 year slices, using taxa data only, (a) using all taxa, and (b) species identifications only.

It is possible that the signals in the more general habitat classes (*Aquatics*, *Meadowland*, *Wood & trees*, *Wetland/marshes* and *Heathland & moorland*) are being smoothed by the variety of sites and environments represented in each time slice. If this is so, then the narrower, and indicator, classes may be of more use in describing important characteristics of time slice environments. The lack of woodland indicator species before 11 ka BP is probably a reliable indication of the start of woodland development after the Younger Dryas. The presence of various kinds of dung indications throughout the later part of the timescale, along with openness indications such as *Meadowland* and *Disturbed/arable* habitats could lend support to the Vera hypothesis of openness in mid-Holocene forests (Vera, 2000; Hodder *et al.*, 2005). Further site based research is necessary before this can be evaluated in greater detail, as recommended by Buckland [*et al.*] (2005).

The standardization method used in the production of BugStats EcoFigs was designed specifically to remove the effect of variable numbers of species or individuals between samples. Although not a rigorous test of the success of this method, the calculation of a correlation coefficient of individual habitat class results against either sample number of species or abundance sums should provide an indication as to whether this has been achieved. Only number of taxa have been used in this example (Figure 6.32), although as a relationship has been established between the number of taxa and number of individuals (Figure 6.31b), the results should be similar. If the representation of habitats produced by standardized BugStats outputs was proportional to the number of species, we could expect a correlation between the habitat class values for each sample and the total number of species per sample. As can be seen from Table 6.19, the raw, untransformed values generally display this relationship (higher R^2 values), with the exception of a few categories which have very low counts. Standardization lowers the correlation significantly in all but a few cases, suggesting that the standardization method successfully counters the affect of varying sample sizes.

A similar effect can be seen when log transformations are applied to the raw data and sums, although the effect is less pronounced in most cases, and even appears to have little effect in the case of the *Standing water* and *Mould beetle* habitats. The significance of this is difficult to interpret, and although the transformation option is not recommended, it may be introducing additional artefacts into the equation due to biases in the coding system.

Table 6.19. The effect of standardization (Std) on the relationship between *sample habitat sums* and *the numbers of species in samples* (species level identifications only). Values are R^2 , the Pearson product moment correlation coefficient, indicating the proportion of the variance in the habitat class values that is attributable to variance in the numbers of species. Values are shaded by magnitude, and a higher value indicates greater correlation.

		Aquatics	Indicators: Standing water	Indicators: Running water	Pasture/Dung	Meadowland	Wood and trees	Indicators: Deciduous	Indicators: Coniferous	Wetlands/marshes	Open wet habitats	Disturbed/arable	Sandy/dry disturbed/arable	Dung/foul habitats	Carrion	Indicators: Dung	Mould beetles	General synanthropic	Dry dead wood	Heathland & moorland	Halotolerant
No transformation																					
NSpec	Raw	0.94	0.90	0.54	0.61	0.88	0.67	0.39	0.27	0.94	0.52	0.63	0.67	0.81	0.71	0.37	0.75	0.66	0.27	0.89	0.59
	Std	0.13	0.79	0.03	0.08	0.18	0.31	0.29	0.00	0.23	0.02	0.17	0.00	0.28	0.04	0.04	0.67	0.49	0.05	0.03	0.07
Ln transformed NSpec and habitat values (Transform abundance Ln(n+1) option in BugStats)																					
NSpec	Raw	0.96	0.72	0.63	0.68	0.90	0.89	0.38	0.20	0.96	0.53	0.81	0.79	0.77	0.52	0.30	0.57	0.56	0.28	0.94	0.57
	Std	0.29	0.70	0.40	0.42	0.46	0.58	0.36	0.07	0.32	0.03	0.67	0.11	0.63	0.30	0.18	0.57	0.57	0.21	0.05	0.38

6.7.6.1 Relationships between reconstructed temperatures and habitats

In the Koivula *et al.* (2004) case study above (section 6.1.6), a slightly wider span of temperature reconstructions were obtained from the forest habitat than the from the farmland fauna. It was therefore suggested that there may be a relationship between the habitats of the species used in temperature reconstructions and the MCR values obtained from them, which might lead to slightly different temperature reconstructions within a region. This could mean that if the dominant habitat of the area represented by a sequence of samples changes with time, then the temperatures reconstructed from different samples could under- or overestimate periods of the past climate. This could be considered, to a certain degree, to relate to the scepticism expressed by Andersen (1993) on the validity of the MCR method. Although the MCR method has been validated with modern data (Atkinson *et al.*, 1987; Coope & Lemdahl, 1996), this has not been done in combination with quantitative reconstruction of the habitats represented by the faunas. The 20 ka dataset was used to test whether any such relationships, between the reconstructed temperatures and the strength of the habitat signals from species level identifications (Figure 6.32b), could be identified for the past 20 000 years using the bulked 1 000 year time slices. Linear regression was used, and although this is not a particularly robust test, or a thorough analysis of the potential relationships, it may allow us to assess the relative amount of variability in each habitat group that can be linearly related to the reconstructed temperatures. Due to the broad nature of the dataset, it may be unwise to attempt more thorough analyses than those presented here, and a more rigorous, site by site analysis would most likely prove more fruitful should this experiment be taken further. The R^2 values for the regression of each habitat group with each reconstructed temperature variable (TValues) are presented in Table 6.20. The strength of the relationships are not given, as the author does not believe that the results would have any ecological grounding in this case, but the direction of the relationship is indicated by the underlining of negative trends. Note that these results do not in any way imply cause and effect, which is liable to be a complex issue, but could form the starting point for further investigations into the relationship between reconstructed habitats and temperatures.

Were the pattern seen in the results from the Koivula *et al.* (2004) data common to the 20 ka dataset, then one would perhaps expect to see a large part of the variation in either the *Wood & trees*, *Sandy/dry disturbed/arable* and/or possibly *Disturbed/arable* habitat groups ‘explained’ by the TMaxDiff and TMinDiff variables. These are the habitat groups that displayed the largest signal differences in the BugStats results for the Koivula *et al.* data (see Figure 6.3), and are thus most likely to be those responding to forest/grassland differences in that context. The low R^2 values for each of these (highlighted in **bold** in Table 6.20) suggest that the pattern is not reproduced in the 20 ka time slices. There are, however, a number of more significant correlations, the most prominent being *Open wet habitats*, *Sandy/dry disturbed/arable* and *Heathland & moorland* (independently) with TMax and TMin. As might be expected, these suggest that these types of environment are more frequent during colder periods. In addition, *Dung/foul habitats* appear to be correlated with TMax and to a lesser extent TMinHi. This could, at least theoretically, be explained in terms of warmer summer temperatures leading to generally higher numbers of herbivores, in a complex relationship involving biological productivity and the ability of the environment to support more large vertebrates (e.g. Brown & Lomolino, 1998).

It should be remembered that only a subset of the species used for habitat reconstruction are used for climate reconstruction, and one should perhaps, ideally, run BugStats on just the MCR species to reduce the risk of coincidental patterns. As mentioned above, the 20 ka dataset only allows the investigation of geographically and chronologically broad scale changes, and the differential representation of environments in the time slices makes the interpretation of statistics difficult. The simple analyses presented here, however, demonstrate that the use of the BugStats system together with an enhanced MCR facility may be extremely useful in increasing our understanding of not only past environmental and climate change, but also the biogeography and ecology of insect populations in general.

Table 6.20. Amount of variance per habitat group explained by the different temperature values output from BugsMCR. Cell values are R², and underlined numbers indicate negative relationships. TMaxDiff and TMinDiff are the TMax and TMin spans respectively (e.g. TMaxHi-TMaxLo). Cells are shaded by R² value, see the text for explanation of the **bold** highlighted cells.

TValue	Aquatics	Indicators: Standing water	Indicators: Running water	Pasture/Dung	Meadowland	Wood and trees	Indicators: Deciduous	Indicators: Coniferous	Wetlands/marshes	Open wet habitats	Disturbed/arable	Sandy/dry disturbed/arable	Dung/foul habitats	Carrion	Indicators: Dung	Mould beetles	General synanthropic	Stored grain pest	Dry dead wood	Heathland & moorland	Halotolerant
TMaxLo	<u>0.09</u>	0.42	0.01	0.20	0.00	0.27	0.17	<u>0.03</u>	0.19	<u>0.66</u>	0.07	<u>0.45</u>	0.60	<u>0.07</u>	0.00	0.34	0.30	0.07	0.01	<u>0.66</u>	0.14
TMaxHi	<u>0.13</u>	0.32	<u>0.00</u>	0.33	<u>0.01</u>	0.29	0.15	<u>0.00</u>	0.19	<u>0.64</u>	0.04	<u>0.46</u>	0.75	<u>0.08</u>	0.02	0.32	0.31	0.04	0.00	<u>0.57</u>	0.08
TMinLo	<u>0.03</u>	0.37	0.11	0.02	0.14	0.34	0.24	<u>0.01</u>	0.03	<u>0.55</u>	0.16	<u>0.43</u>	0.18	0.01	0.00	0.31	0.21	0.06	0.17	<u>0.77</u>	0.09
TMinHi	<u>0.03</u>	0.13	0.01	0.34	<u>0.08</u>	0.58	0.45	0.03	0.07	<u>0.59</u>	0.02	<u>0.55</u>	0.51	<u>0.16</u>	0.09	0.25	0.18	0.01	0.15	<u>0.55</u>	0.00
TRangeLo	<u>0.00</u>	0.01	<u>0.01</u>	<u>0.08</u>	0.09	0.37	<u>0.36</u>	0.10	0.00	0.11	0.00	0.17	<u>0.03</u>	0.06	<u>0.07</u>	<u>0.02</u>	<u>0.00</u>	0.00	<u>0.32</u>	0.11	0.09
TRangeHi	0.00	<u>0.23</u>	<u>0.19</u>	0.00	<u>0.22</u>	<u>0.27</u>	<u>0.23</u>	0.01	0.00	0.30	<u>0.17</u>	0.25	<u>0.02</u>	<u>0.05</u>	<u>0.00</u>	<u>0.21</u>	<u>0.11</u>	0.04	<u>0.32</u>	0.55	0.03
TMaxDiff	<u>0.01</u>	<u>0.27</u>	<u>0.27</u>	0.11	0.16	0.01	<u>0.03</u>	0.25	<u>0.01</u>	0.08	0.10	0.03	0.01	<u>0.00</u>	0.11	<u>0.07</u>	<u>0.02</u>	<u>0.11</u>	<u>0.16</u>	0.19	<u>0.18</u>
TMinDiff	0.01	<u>0.26</u>	<u>0.13</u>	0.05	<u>0.47</u>	0.03	<u>0.02</u>	0.08	<u>0.00</u>	0.15	0.16	0.09	<u>0.00</u>	<u>0.14</u>	0.02	<u>0.12</u>	<u>0.07</u>	0.06	0.06	0.32	<u>0.15</u>

6.7.7 Conclusions

The application of spatial statistics, or geostatistics, to time slices extracted from BugsCEP data would require careful time-depth modelling of each site to be of real research value. The time slices used in here are simply bulked date categories, with limited regard for the errors involved, and no attempt to interpolate or extrapolate sample dates through sequences. To do so without specific reference to the original publications would be risky to say the least. This has important consequences for the reconstructions in that they should be regarded as extremely tenuous conclusions at best. In reality, some of the inferences drawn may have been misleading due to dating issues. Palaeoenvironmental science is not unique in being subject to these problems, and indeed practically any work which attempts to utilize data from a large spatial and chronological span, both in terms of sample dates and when the research was undertaken, will have similar problems.

Although the examples above use artificially compounded assemblages, they serve to illustrate how careful examination of the jackknifed results, along with the climate space maps for a sample, can help resolve difficulties in the interpretation of MCR climate reconstructions. The use of BugStats illustrates that, although the interpretation of samples which enclose varying environments, either by virtue of their resolution/size, or in this case geographical extension, is problematic, the tools offered by BugsCEP can help considerably. The standardization method (%SumRep) employed in the construction of EcoFigs appears to perform well in this situation, and effectively compensates for the variations in species numbers that are common in sequences of palaeontomological samples. Further testing, however, at the site level, with both modern and fossil assemblages is required to further validate the method.

7 Conclusions

As the only available software package for the storage and analysis of fossil beetle data, with modern reference data included, it is difficult to compare BugsCEP with similar programs. The only alternative to using BugsCEP is the manual construction of species lists, ideally in spreadsheets, using checklists such as Gustafsson (2005) to provide the taxonomic order and synonymy. Ecology and distribution data can then be compiled for these lists by the use of published and unpublished literature and various Internet resources, the latter only recently becoming partly feasible, and the former requiring good library and museum collection access. Ecological classification of the species is often undertaken on the basis of specialist knowledge, although without a database it is difficult to be consistent between projects. The construction of diagrams similar to the Bugs EcoFig habitat reconstruction ones is laborious without semi-automated software assistance, and climate reconstruction through previous MS-DOS MCR packages, or by hand, is cumbersome. In addition to these improvements in efficiency, BugsCEP also provides a number of advanced search and query tools, such as rapidly obtaining the known fossil record for a species, which is extremely difficult to duplicate without the database.

The primary aims of the BugsCEP project have been fulfilled in this thesis, in terms of the restructuring of the database, and the implementation of climate and habitat reconstruction systems. As always, there is room for improvement, as there will always be with research orientated software, and a number of possibilities have been discussed throughout the thesis. In particular, some desirable functions are missing from this release of the software, the lack of Geographical Information System (GIS) functionality being perhaps the most important. Beetle distribution data are prime candidates for a geographical database, as are the location of fossil sites. The integration of point distribution data, climate data, and chronologically delimited fossil data would provide the foundations for an extremely useful investigative and modelling system. This is one of the primary directions for the future development of BugsCEP, and the research possibilities for GIS interrogation of the data are enormous. Before venturing into the world of GIS, however, one must be reminded of the data quality issues involved in drawing biogeographical conclusions from variably dated data. That is to say, one should be fully aware of the quality of dating evidence for every site that is involved in multi-site inferences, and the blanket application age depth models may confuse the results. One should also be aware of the implications of spatial variation of species presence between sites and dates. Taphonomic issues render 'absence' a very debatable concept – is a species not found because it did not live there, or because the sample size was too small, or because the sampling location did not provide a good trap for that species? These questions of sample representativeness are very hard to assess, and require a working knowledge of the ecology of the individual species and the depositional and preservational environment of the site. Suffice it say that multi-proxy analyses can be of great benefit in such situations – the additional proxies providing support information for the insect interpretations in terms of extra explanatory variables, and hopefully providing proxy information for some of the issues outlined above. BugsCEP provides an indication of the availability of that data, and could act as a starting point for assessing the availability of existing sites most suitable for detailed multi-proxy studies.

Little attempt has been made to address specific preservational and taphonomic issues here, and a detailed appraisal of the ecology of all species mentioned is beyond the scope of the thesis. The aim was more to provide some starting points for future research than to fully explore all angles in detail. BugsCEP includes a mixture of archaeological and other Quaternary sites, and it is important to be very careful when mixing these in large scale GIS analyses due to both the possible implications of human constructed environments, or human impact on the landscape, and the inevitable increased variability of preservation seen at archaeological sites. One could argue that the two groups of samples are not comparable (due to differences in sampling aims, standards and taphonomy), but as long as standard sampling and processing methodology has been followed then the identified faunas should be comparable, as long as the researcher is aware of the potential problems involved.

The successful implementation of BugStats, and the development of the Bugs EcoCode habitat classification system, is only the start of the possibilities for semi-quantitative environmental reconstruction from fossil beetle remains from the BugsCEP database. There are a number of theoretical issues that need to be investigated before the Bugs EcoFig system can be considered universally applicable. The primary concerns are, however, those that are common to any (semi-)quantitative reconstruction method, such as sample representativeness in terms of time exposure/collection, catchment complexity, the abundance and richness of species in the catchment habitats, deposit type (well, bog, pool, etc.), and the appropriateness of the numerical methods. These issues aside, the system does produce useful results which are comparable with those published independently by other researchers. Further testing, by a wider group of users, and on a wider range of contexts, especially modern references sites, is necessary for full validation of the method.

At the onset of this project, the intention was to create a system for assisting the refinement of species thermal envelopes, and even improve on the existing reference dataset. Although a prototype system was developed, using a combination of MS Access and ESRI ArcView 3.1 (Buckland & Buckland, 2002), the task proved far too time consuming to begin enhancing the envelopes within the scope of this project. The implementation of existing MCR dataset (Atkinson *et al.*, 1986), and the development of a new form of the temperature reconstruction software integrated with the BugsCEP site database proved successful and reliable. It also provided the basis for enhancements to the method which may prove suitable for refining the climate reconstructions made from fossil beetles.

7.1 Wider Applications

The environmental and climatic sensitivity of insects makes them particularly useful in studying the Arctic interface zones that people have struggled to survive in for millennia. These regions not only provide challenges for the human populations attempting to survive there, but also tend to be more susceptible to the impacts that these attempts bring upon them. Beyond the reaches of historical documentation, the impact of northern peoples is poorly understood, and there is great scope its the study through proxy indicators – which are the only alternative available for the reconstruction of the past, which cannot be measured directly.

Apart from potentially containing a proxy record of human impact at numerous sites throughout Europe, BugsCEP has the potential to help in the prediction of future human and natural impacts on the environment by way of predictive modelling. The BugsMCR ‘Predict’ module is a start, and the refinement of this system, in combination with fine-tuning and testing of the BugStats reconstruction component, could provide a base for modelling the potential changes in insect populations in the future. In the areas of forestry and agriculture, rewilding/naturalization projects, and the development of sustainable resource management there are great potentials for changes in the insect populations that could have far reaching consequences for both the biodiversity and industrial or agricultural sustainability of the regions concerned. BugsCEP’s database of Quaternary sites contains a wealth of palaeo-biodiversity information that would be of benefit in any attempt to model the present or future changes in beetle populations, especially under constrained change conditions.

7.2 Databasing the Humanities

Although it is a multi disciplinary database, and not a humanities database in the strict sense, BugsCEP can be used to provide support for integrated studies into the history of the interactions between people and the environment. This is a major theme in current research at the European level, and includes such diverse aspects as predicting the public effects of sustainable development policies to climate change studies. The humanities appear to be entering a paradigm where empirical data, and thus the large scale management of this, is essential for underpinning conclusions drawn on the human condition. This is especially important in the eyes of the public, and other scientific disciplines, such as the environmental sciences, where theories of the economic rationality of resource use alone cannot satisfactorily explain patterns of past impacts. As yet, the large scale databasing, and even

computerization, of the humanities is underdeveloped in Sweden (Vetenskapsrådet, 2005), with the notable exception of facilities such as the Demographic Data Base (Demografiska databasenⁱ), a number of linguistic databases, and projects associated with the ‘humanities IT environments’ (e.g. HUMlabⁱⁱ). There are a number of large palaeoenvironmental databases (see Chapter 1) which could be usefully integrated with archaeological, and related, data, were it available in a transferable form. The human aspect is frequently missing in palaeoenvironmental studies from the ecological and environmental sciences, often due to a lack of access to the data or the skills to interpret it. The complex nature of archaeological dating, and the wealth of data confined to so called grey literature (e.g. site reports, consultancy work), are perhaps two of the main reasons why raw archaeological site data are not currently available, on a large scale, at a suitable resolution for such integration. Database initiatives, such as SEAD (The Strategic Environmental Archaeology Database, Buckland *et al.*, 2006), and hopefully others sponsored by the EU’s research infrastructure initiativesⁱⁱⁱ, should help to bridge this gap and provide greater scope for inter-faculty cooperation. In addition, interdisciplinary cooperation, and the use of inherently non-disciplinary resources such as HUMlab could provide innovative solutions to data mining and visualization problems that would not normally be thought of within the confines of traditional disciplinary boundaries.

7.3 Future Directions and Final Thoughts

This thesis may ask more questions than it answers, as is often the case in doctoral research. Whilst BugsCEP provides the basis for improving the efficiency and power of palaeoenvironmental investigations, there are many aspect which could be developed further. The expansion of the scope of the database, to include more taxa, including Diptera, Formicidae and Trichoptera, which are amongst the groups more commonly found fossil, is a high priority. In addition, the start that was made on the parallel Bugs-like molluscan database, Slugs, which was begun with the late David Keen, should be built upon. The geographical expansion of the database, into southern and eastern Europe, but also other areas of the world would be of great benefit to the Quaternary research community, and work is already in progress to assemble a North American version. The Egyptian version, EgBugs (Buckland *et al.*, *in press*), is proof of concept for the viability of this idea. There is also potential for the growth of the use of Coleoptera in climate change studies, with the improvement and refinement of the MCR dataset.

Besides these grand schemes, there are also a number of smaller enhancements that could improve the usefulness of BugsCEP, such as time-scaled graphing of outputs, additional variables and statistics, and possibly even the calculation of age-depth curves. The unfinished, experimental systems, such as the MCR jackknifing, and querying by fossil date routines, need to be completed, tested and released. Further testing of the Bugs EcoCode system, including comparison with other existing habitat classifications on both modern and fossil faunas is also necessary.

The development of a web based version of BugsCEP, including GIS and data submission facilities, would be an important step towards providing a system which can be integrated into a resource for the multi-proxy interrogation of Quaternary data. Whilst a prototype structure was created using MS SQL-Server (primarily to enable the construction of the database diagrams used in this thesis), the data transfer and web development aspects are yet to be undertaken, and are a high priority for future development. Similarly important is the inclusion of illustrations, which will undoubtedly increase the public appeal and teaching potential.

Although BugsCEP can be used for complex environmental and climatic reconstructions, it only uses insect data, and there is always a need for comparison with other proxy sources. Multi-proxy databases, such as LNEED and SEAD (see Chapter 1) are essential if we are to improve on our ability to understand the past. Single proxy databases, such as BugsCEP, and others mentioned in Chapter 1,

ⁱ <http://www.ddb.umu.se/>

ⁱⁱ <http://www.humlab.umu.se/humlabinenglish>

ⁱⁱⁱ <http://cordis.europa.eu/esfri/>

are vital building blocks in these developments. These databases, and their associated web portals and support networks should strive to reach potential users outside of the traditional educational establishments, and the immediately relevant disciplines of Quaternary geology and archaeology, where a diversity of backgrounds may bring new approaches to the interpretation of the data.

8 References

- Aaby, B. & Berglund, B.E. (1986). Characterization of peat and lake deposits. In B.E. Berglund (ed.), *Handbook of Holocene Palaeoecology and Palaeohydrology*. J. Wiley & Son, Chichester, 851-858.
- Alley, R.B., Meese, D.A., Shuman, C.A., Gow, A.J., Taylor, K.C., Grootes, P.M., White, J.W.C., Ram, M., Waddington, E.D., Mayewski, P. & Zielinski, G.A. (1993). Abrupt increase in the Greenland snow accumulation at the end of the Younger Dryas event. *Nature* **362**, 527-529.
- Andersen, T., Ligaard, S., Pedersen, T. & Soli, G.E.E. (1990). Pitfall catches of Carabidae and Staphylinidae (Coleoptera) in a temporary protected forest area on the Eidanger Peninsula, Telemark, SE Norway. *Fauna Norvegica* **B37**, 13-22.
- Andersen, J. (1993). Beetle remains as indicators of the climate in the Quaternary. *Journal of Biogeography* **20**, 557-562.
- Andersen, J. (1996). Do beetle remains reliably reflect the macroclimate in the past? - a reply to Coope & Lemdahl. *Journal of Biogeography* **23**, 120-121.
- Ashworth, A.C. (1972). A Late-glacial insect fauna from Red Moss, Lancashire, England. *Entomologica Scandinavia* **3**, 211-224.
- Ashworth, A.C. (2001). Chapter 8: Perspectives on Quaternary Beetles and Climate Change. In L.C. Gerhard, W.E. Harrison, and B.M. Hanson (eds.), *Geological Perspectives of Global Climate Change*. American Association of Petroleum Geologists Studies in Geology **47**, Tulsa, Oklahoma, 153-168.
- Ashworth, A.C., Buckland, P.C. & Sadler, J.P. (eds.) (1997) *Studies in Quaternary Entomology - An Inordinate Fondness for Insects*. *Quaternary Proceedings* **5**.
- Atkinson, T.C., Briffa, K.R., Coope, G.R., Joachim, J.M. & Perry, D.W. (1986). Climatic calibration of coleopteran data. In B. E. Berglund (ed.), *Handbook of Holocene Palaeoecology and Palaeohydrology*. J. Wiley & Son, Chichester, 851-858.
- Atkinson, T.C., Briffa, K.R. & Coope, G.R. (1987). Seasonal temperatures in Britain during the past 22,000 years, reconstructed using beetle remains. *Nature* **325**, 587-592.
- Barnekow, L. (1999a). *Holocene vegetation dynamics and climate changes in the Torneträsk area, northern Sweden*. PhD, Quaternary Geology, Lund, Sweden.
- Barnekow, L. (1999b). Holocene treeline dynamics and inferred climatic changes in the Abisko area, northern Sweden, based on macrofossil and pollen records. *The Holocene* **9**, 253-265.
- Battarbee, R.W., Gasse F. & Stickley, C.E. (eds.) (2004). *Past Climate Variability through Europe and Africa* **6**. Springer, Dordrecht.
- Beanlieu, J-L. de & Reille, M. (1992). The last climatic cycle at la Grande Pile (Vosges, France): a new pollen profile. *Quaternary Science Reviews* **11**, 431-438.
- Bell, M. & Walker M.J.C. (2005). *Late Quaternary Environmental Change - Physical and Human Perspectives* (2nd ed.). Longman, Essex.
- Bennett, K.D. (2002). Comment: the Greenland 8200 cal. yr BP event detected in loss-on-ignition profiles in Norwegian lacustrine sediment sequences, A. Nesje and S.O. Dahl (2001). *Journal of Quaternary Science* **17**, 97-99.
- Berglund, B.E. & Ralska-Jasiewiczowa, M. (1986). Pollen analysis and pollen diagrams. In B.E. Berglund (ed.), *Handbook of Holocene palaeoecology and palaeohydrology*. Wiley & Sons, Chichester, 455-484.
- Bergman, J., Hammarlund, D., Hannon, G., Barnekowa, L. & Wohlfarth, B. (2005). Deglacial vegetation succession and Holocene tree-limit dynamics in the Scandes Mountains, west-central Sweden: stratigraphic data compared to megafossil evidence. *Review of Palaeobotany and Palynology* **134**, 129-151.
- Bigler C., Larocque I., Peglar S.M., Birks H.J.B. & Hall R.I. (2002). Quantitative multiproxy assessment of long-term patterns of Holocene environmental change from a small lake near Abisko, northern Sweden. *The Holocene* **12**, 481-496.

- Bigler, C., Grahn, E., Larocque, I., Jeziorski, A. & Hall, R.I. (2003). Holocene environmental change at Lake Njulla (999 m a.s.l.), northern Sweden: a comparison with four nearby lakes along an altitudinal gradient. *Journal of Paleolimnology* **29**, 13-29.
- Birks, H.J.B. (1981). The use of pollen analysis in the reconstruction of past climates: a review. In T.M.L. Wigley, M.J. Ingram, & G. Farmer (eds.), *Climate and history*. Cambridge: Cambridge University Press, 1-38.
- Birks, H.J.B. (1995). Quantitative palaeoenvironmental reconstructions. In D. Maddy & J.S. Brew (eds.), *Statistical modelling of Quaternary science data*. Technical Guide **5**, Quaternary Research Association, Cambridge, UK, 161–254.
- Birks, H.J.B. (2003). Quantitative palaeoenvironmental reconstructions from Holocene biological data. In A. Mackay, R.W. Battarbee, H.J.B. Birks & F. Oldfield (eds.), *Global Change in the Holocene*. Arnold, London, 107-123.
- Birks, H.J.B. & Line, J.M. (1992). The use of rarefaction analysis for estimating palynological richness from Quaternary pollen-analytical data. *The Holocene* **2**, 1-10.
- Box, G.E.P. & Draper, N.R. (1987). *Empirical Model-Building and Response Surfaces*. Wiley.
- Bray, P.J., Blockley, S.P.E., Coope, G.R., Dadswell, L.F., Elias, S.A., Lowe, J.J. & Pollard, A.M. (2006). Refining mutual climatic range (MCR) quantitative estimates of palaeotemperature using ubiquity analysis. *Quaternary Science Reviews* **25**, 1865-1876.
- Bronk Ramsey, C. (1995). Radiocarbon calibration and analysis of stratigraphy: The OxCal program, *Radiocarbon* **37** (2), 425-430
- Broström, A. (2002). *Estimating source area of pollen and pollen productivity in cultural landscapes of southern Sweden - developing a palynological tool for quantifying past plant cover*. LUNDQUA Thesis **46**, Lund University, Lund.
- Buckland, P.C. (1975). Synanthropy and the deathwatch : a discussion. *Naturalist* **100** (933). Hull, 37-42.
- Buckland, P.C. (1981). Insect remains from beneath the Brigg 'Raft'. In S. McGrail (ed.), *The Brigg 'Raft' and her prehistoric environment*. British Archaeological Reports **89**. Oxford, 155-175.
- Buckland, P.C. (1982). The cover sands of North Lincolnshire and the Vale of York. In B.H. Adlam, C.R. Fenn & L. Morris (eds.), *Papers in Earth Studies*. Geobooks, Norwich, 143-178.
- Buckland, P.C. (1990). Granaries, Stores and Insects. The Archaeology of Insect Synanthropy. In D. Fournier & F. Sigaut (eds.), *La préparation alimentaire des céréales*,. Rapports présentés à la Table ronde, Ravello au Centre Universitaire pour les Biens culturels, avril 1988. PACT, Rixensart, Belgium, 69-81.
- Buckland, P.C. [Assisted by Buckland, P.I. & Hughes, D.] (2005). Palaeoecological evidence for the Vera hypothesis? In K.H. Hodder, J.M. Bullock, P.C. Buckland & K.J. Kirby (eds.), *Large herbivores in the wildwood and modern naturalistic grazing systems*. English Nature Research Reports **648**, 62-116.
- Buckland, P.C., Amorosi, T., Barlow, L.K., Dugmore, A.J., Mayewski, P.A., McGovern, T.H., Ogilvie, A.E.J., Sadler, J.P. & Skidmore, P. (1996). Bioarchaeological and climatological evidence for the fate of Norse farmers in medieval Greenland. *Antiquity* **70**, 88-96.
- Buckland, P.C., Buckland, P.I. & Skidmore, P. (1998). Insect remains from GUS: an interim report. In J. Arneborg & H.C. Gulløv (eds.), *Man, Culture and Environment in Ancient Greenland*. Danish National Museum & Danish Polar Centre, Copenhagen, 74-79.
- Buckland, P.C., Sadler, J. P. & Smith, D. (1993). An Insect's Eye-view of the Norse Farm. In C.E. Batey, J. Jesch & C.D. Morris (eds.), *The Viking Age in Caithness, Orkney and the North Atlantic*. Edinburgh University Press, UK, 518-528.
- Buckland, P.I. (2000). *An introduction to palaeoentomology in archaeology and the BUGS database management system*. Magister paper, Archaeology and Sámi Studies, University of Umeå, Sweden. Available online: <http://www.umu.se/archaeology/publikation/uppsatser/cd/Buckland2000/Buckland2000.pdf>

- Buckland, P.I. & Buckland, P.C. (2002). How can a database full of Bugs help reconstruct the climate? In G. Burenhult & J. Arvidsson (eds.), *Archaeological informatics: pushing the envelope. CAA 2001. Computer applications and quantitative methods in archaeology. Proceedings of the 29th conference, Gotland, April 2001*, British Archaeological Reports International Series **1016**. Oxford, 453-462.
- Buckland, P.I., Buckland, P.C., Panagiotakopulu, E. & Sadler, J.P. (*in press*). A Database for Egyptian Entomology. *The Journal of Egyptian Entomology*.
- Buckland, P.I., Olofsson, J. & Engelmark, R. (2006). SEAD – Strategic Environmental Archaeology Database - Planning report. *MAL Reports* **2006-31**. (Unpublished report).
- Buckland, P.I., Yuan Zhuo, D. & Buckland, P.C. (1997). Towards an expert system in Palaeoentomology. In A.C. Ashworth, P.C. Buckland & J.P. Sadler (eds.), *Studies in Quaternary Entomology - An Inordinate Fondness for Insects. Quaternary Proceedings* **5**, 67-78.
- Böhme, J. (2005). *Die Käfer Mitteleuropas. K. Katalog (Faunistische Übersicht)* (2nd ed.). Spektrum Academic, Munich.
- Camilloni, I. & Barros, V. (1997). On the Urban Heat Island Effect Dependence on Temperature Trends. *Climatic Change* **37** (4), 665-681.
- Clarke, T.E., Levin, D.B., Kavanaugh, D.H., & Reimchen T.E. (2001). Rapid evolution in the *Nebria gregaria* group (Coleoptera: Carabidae) and the paleogeography of the Queen Charlotte Islands. *Evolution* **55**, 1408–1418.
- Codd, E.F. (1970). A relational model of data for large shared data banks. *Communications of the ACM* **13** (6), 377–387. Retrieved from ACM online, November 30, 2006: <http://www.acm.org/classics/nov95/toc.html>
- Colinvaux, P.A. (1973). *Introduction to ecology*. Wiley, New York.
- Cong, S. & Ashworth, A.C. (1997). The use of Correspondence Analysis in the analysis of fossil beetle assemblages. In A.C. Ashworth, P.C. Buckland & J.P. Sadler (eds.), *Studies in Quaternary Entomology - An Inordinate Fondness for Insects. Quaternary Proceedings* **5**, 79-82.
- Coope, G.R. (1959). A Late Pleistocene insect fauna from Chelford, Cheshire. *Proceedings of the Royal Society of London* **B151**, 70-86.
- Coope, G.R. (1978). Constancy of insect species versus inconstancy of Quaternary environments. In L.A. Mound & N. Waloff (eds.), *Diversity of Insect Faunas*. Symposia of the Royal Entomological Society of London **9**. Blackwell, Oxford, 176-187.
- Coope, G.R. & Böcher, J. (2000). Coleoptera from the Late Weichselian deposits at Nørre Lyngby, Denmark and their bearing on palaeoecology, biogeography and palaeoclimate. *Boreas* **29**, 26-34.
- Coope, G.R. & Joachim, M.J. (1980). Lateglacial environmental changes interpreted from fossil Coleoptera from St. Bees, Cumbria, N.W. England. In J.J. Lowe, J.M. Gray & J.E. Robinson (eds.), *Studies in the Lateglacial of North-West Europe*. Pergamon, Oxford, 55-68.
- Coope, G.R. & Lemdahl, G. (1995). Regional differences in the Lateglacial climate of northern Europe based on coleopteran analysis. *Journal of Quaternary Science* **10**, 391-395.
- Coope, G.R. & Lemdahl, G. (1996). Validations for the use of beetle remains as reliable indicators of Quaternary climates: a reply to the criticisms by Johan Andersen. *Journal of Biogeography* **23**, 115-120.
- Coope, G.R., Lemdahl, G., Lowe, J.J. & Walkling, A. (1998). Temperature gradients in northern Europe during the last glacial-Holocene transition (14-9 ¹⁴C kyr BP) interpreted from coleopteran assemblages. *Journal of Quaternary Science* **13**, 419-434.
- Coope, G.R., Shotton, F.W. & Strachan, I. (1961). A Late Pleistocene fauna and flora from Upton Warren, Worcestershire. *Philosophical Transactions of the Royal Society of London* **B244**, 379-421.

- Davis, A.L.V., Scholtz, C.H. & Philips, T.K. (2002). Historical biogeography of scarabaeine dung beetles. *Journal of Biogeography* **29**, 1217-1256.
- Davis, A.L.V., van Aarde, R.J., Scholtz, C.H. & Delpont, J.H. (2002). Increasing representation of localized dung beetles across a chronosequence of regenerating vegetation and natural dune forest in South Africa. *Global Ecology & Biogeography* **11**, 191-209.
- Dockerty, T. & Lovett, A. (2003). A Location-centred, GIS-based Methodology for Estimating the Potential Impacts of Climate Change on Nature Reserves. *Transactions in GIS* **7**, 345-370.
- Drotz, M.K. (2003). *Population genetics and phylogeography of the West Palearctic Agabus bipustulatus diving beetle complex*. PhD thesis, Molecular Biology, Umeå University, Sweden.
- Duff, A. (1993). *Beetles of Somerset: their status and distribution*. Somerset Archaeological & Natural History Society, Taunton.
- Edwards, K.J., Langdon, P.G. & Sugden, H. (2006). Separating climatic and possible human impacts in the early Holocene: biotic response around the time of the 8200 cal. yr BP event. *Journal of Quaternary Science* **22** (1), 77-84.
- Elias, S.A. (1994). *Quaternary Insects and Their Environments*. Smithsonian Institution Press, Washington.
- Eliasson, L. & Kishonti, I. (2003). *Rapport över arkeologiska slutundersökning: Öresundsförbindelsen, Lockarp 7B*. Rapport Nr 17, Malmö Kulturmiljö.
- Engelmark, R. (1996). North Sweden. In: B.E. Berglund, H.J.B. Birks & M. Ralska-Jasiewiczowa (eds), *Palaeoecological events during the last 15000 years: regional syntheses of palaeoecological studies of lakes and mires in Europe*. Wiley, Chichester, 266-276.
- Engelmark, R. & Linderholm, J. (in prep.). Människa och landskap – en komplicerad dynamik. Miljöarkeologi inom Öresundsförbindelsen.
- Eyre, M.D., Foster, G.N., Luff, M.L. & Rushton, S.P. (2006). The Definition of British Water Beetle Species Pools (Coleoptera) and their Relationship to Altitude, Temperature, Precipitation and Land Cover Variables. *Hydrobiologia* **560**, 121-131.
- Eyre, M.D. & Luff, M.L. (1990) A preliminary classification of European grassland habitats using carabid beetles. In N.E. Stork (ed.), *Ground Beetles: Their Role in Ecological and Environmental Studies*. Intercept, Andover, 227-236.
- Fauna Entomologica Scandinavica. Series (various publishers). ISSN: 0106-8377.
- Foster, G. (2000). *A review of the scarce and threatened Coleoptera of Great Britain. Part 3. Water Beetles*. UK Joint Nature Conservation Committee, Peterborough.
- Geiger, R., Aron R.H. & Todhunter, P. (2003). *The Climate near the ground* (6th ed.). Rowman & Littlefield Publishers, Lanham, MD, USA.
- Gifford D. (ed.) (1996). *Access 95 Unleashed*. Sams Publishing, Indianapolis.
- Good, J.A. & Giller, P.S. (1991). The diet of predatory Staphylinid beetles - a review of records. *Entomologist's Monthly Magazine* **127**, 77-89.
- Graham, C.H. & Hijmans, R.J. (2006). A comparison of methods for species ranges and species richness. *Global Ecology & Biogeography* **15**, 578-587.
- Greenwood, M.T., Agnew, M.D. & Wood, P.J. (2003). The use of caddisfly fauna (Insecta: Trichoptera) to characterise the Late-glacial River Trent, England. *Journal of Quaternary Science* **18**, 645-661.
- Greenwood, M.T., Wood, P.J. & Monk, W.A. (2006). The use of fossil caddisfly assemblages in the reconstruction of flow environments from floodplain paleochannels of the River Trent, England. *Journal of Paleolimnology* **35**, 747-761.
- Grimaldi, D. & Engel, M.S. (2005). *Evolution of the insects*. Cambridge University Press, Cambridge.
- Gustafsson, B. (2005). *Catalogus Coleopterorum Sueciae. Electronic form: CATCOL2004.XLS, revised 2005-02-01*. Swedish Museum of Natural History. Downloadable from: <http://www2.nrm.se/en/catalogus.html.se>

- Gärdenfors, U. (2000). *Hur rödlistas arter? Manualer och riktlinjer* [How to red-list species. Manual and guidelines]. ArtDatabanken, SLU, Uppsala [In Swedish, English Summary].
- Hall, A.R. & Kenward, H.K. (1990). *Environmental evidence from the Colonia*. Archaeology of York, **14/6**. Council for British Archaeology for York Archaeological Trust.
- Hall, A.R., Jones, A.K.G., Kenward, H.K. & O'Connor, T.P. (1992). The environmental evidence. In D. Tweddle (ed.), *The Anglian Helmet from Coppergate*. Archaeology of York **17/8**, Council for British Archaeology for York Archaeological Trust, 875-880; 903-907.
- Hammer, Ø. & Harper, D.A.T. (2006). *Paleontological data analysis*. Blackwell Publishing, Malden, MA.
- Hansen, M. (1987). *The Hydrophiloidea (Coleoptera) of Fennoscandia and Denmark*. Fauna Entomologica Scandinavica **18**. Scandinavian Science Press, Leiden.
- Harde, K.W. (1998). *A Field Guide in Colour to Beetles*. Octopus, London.
- Hellqvist, M. & Lemdahl, G. (1996). Insect Assemblages and Local Environment in the Mediaeval Town of Uppsala, Sweden. *Journal of Archaeological Science* **23**, 873-881.
- Henriksen, K.L. (1933). Undersøgelser over Danmark-Skånes kvartære Insektfauna. *Videnskabelige Meddelelser fra Dansk naturhistorisk Forening*, **96**, 77-355.
- Hicks, S. (1993). Pollen evidence of localized impact on the vegetation of northernmost Finland by hunter-gatherers. *Vegetation History and Archaeobotany* **2**, 137-144.
- Hodder, K.H., Bullock, J.M., Buckland P.C. & Kirby, K.J. (2000). Large herbivores in the wildwood and modern naturalistic grazing systems. *English Nature Research Reports* **648**.
- Hodge, P.J. & Jones, R.A. (1995). *New British Beetles. Species not in Joy's practical handbook*. British Entomological and Natural History Society, Reading.
- Holm, L. (1991). *The Use of Stone and Hunting of Reindeer - A Study of Stone Tool Manufacture and Hunting of Large Mammals in the Central Scandes c. 6000-1 BC*. PhD thesis, Archaeology, University of Umeå, Sweden.
- Huppert, A. & Solow, A.R. (2004). Reconstructing climate from fossil beetle assemblages. *Proceedings of the Royal Society of London* **B271**, 1125-1128.
- Hyman, P.S. (1992). *A review of the scarce and threatened Coleoptera of Great Britain*, Part 1 (Revised & updated by M.S.Parsons). UK Joint Nature Conservation Committee, Peterborough.
- IUCN (2001). *IUCN Red List Categories and Criteria: Version 3.1*. IUCN Species Survival Commission. IUCN, Gland, Switzerland and Cambridge, UK.
- Iversen, J. (1944). *Viscum, Hedera and Ilex as climatic indicators. A contribution to the study of past-glacial temperature climate*. *Geologiske Foreningen Förhandlingar* **66**, 463-483.
- James, P.E. (1972). *All Possible Worlds, a history of geographical ideas*. The Odyssey Press, New York.
- Jaques, D., Kenward, H. & Johnstone, C. (2000). Assessment of biological remains from excavations at Hayton, East Riding of Yorkshire (site code KINCM1995.1020). *Reports from the Environmental Archaeology Unit, York*, **2000/35**. York.
- Jessop, L. (1986). *Coleoptera: Scarabaeidae. Handbooks for the Identification of British Insects* **5** (11). Royal Entomological Society of London.
- Joachim, M. (1978). *Late-glacial Coleopteran assemblages from the west coast of the Isle of Man*. Unpubl. Ph.D. thesis, University of Birmingham, UK.
- Jongman, R.H.G., ter Braak, C.J.F. & van Tongeren, O.F.R. (eds.) (1995). *Data analysis in community and landscape ecology*. Cambridge University Press, Cambridge.
- Kenward, H.K. (1975). Pitfalls in the environmental interpretation of insect death assemblages. *Journal of Archaeological Science* **2**, 85-94.
- Kenward, H.K. (1978). The Analysis of Archaeological Insect Assemblages: a New Approach. *Archaeology of York* **19/1**. Council for British Archaeology for York Archaeological Trust, London.

- Kenward, H.K. (1985). Outdoors-indoors? The outdoor component of archaeological insect assemblages. In N.R.J. Fieller, D.D. Gilbertson & N.G.A. Ralph (eds.), *Palaeobiological Investigations: Research Design, Methods and Data Analysis*. Association for Environmental Archaeology Symposium 5B, 97-104.
- Kenward, H.K. (1992). Rapid recording of archaeological insect remains - a reconsideration. *Circaea, The Journal of the Association for Environmental Archaeology* **9**, 81-88.
- Kenward, H.K. (2001). Insect remains from the Romano-British ditch terminal at the Flodden Hill Rectilinear Enclosure. *Reports from the Environmental Archaeology Unit, York* **2001/49**. York.
- Kenward, H.K. & Hall, A.R. (1995). *Biological Evidence from 16-22 Coppergate*. Archaeology of York **14/7**. Council for British Archaeology for York Archaeological Trust, York.
- Kenward, H.K. & Carrott, J. (2001). Insect species associations characterise past occupation sites. *Journal of Archaeological Science* **33**, 1452-1473.
- Knapp, S (2000). What's in a name? *Nature* **408**, 33.
- Koch, K. (1989-92). *Die Käfer Mitteleuropas. Ökologie*, **1-3**. Goecke & Evers, Krefeld.
- Koivula, M., Hyyryläinen, V. & Soininen, E. (2004). Carabid beetles (Coleoptera: Carabidae) at forest-farmland edges in southern Finland. *Journal of Insect Conservation* **8**, 297-309.
- Kovach, W.L. (1995). Multivariate Data Analysis. In D. Maddy & J.S. Brew (eds.), *Statistical Modelling of Quaternary Science Data*. Quaternary Research Association, Technical Guide **5**, 1-38.
- Krebs, C.J. (1999). *Ecological methodology* (2nd ed.). Benjamin/Cummings, Menlo Park, California.
- Kullman, L. (1992). Orbital forcing and tree-limit history: hypothesis and preliminary interpretation of evidence from Swedish Lapland. *The Holocene* **2**, 131-37.
- Kullman, L. (1999). Early Holocene tree growth at a high elevation site in northernmost Scandes of Sweden (Lapland): a palaeobiogeographical case study based on megafossil evidence. *Geografiska Annaler* **81A**, 63-74.
- Landin, B.O. (1961). Ecological Studies on Dung-Beetles. *Opuscula Entomologica, Suppl.* **19**.
- Lassau, S.A., Hochuli, D., Cassis, F.G. & Reid, C.A.M. (2005). Effects of habitat complexity on forest beetle diversity: do functional groups respond consistently? *Diversity and Distributions* **11**, 73-82.
- Lawrence, J.F. & Newton, A.F.Jr. (1995). Families and subfamilies of Coleoptera (with selected genera, notes, references and data on family-group names). In J. Pakaluk & S.A. Slipinski (eds.), *Biology, Phylogeny, and Classification of Coleoptera* **2**. Museum & Institute of Zoology, Warsaw, 779-1008.
- Lemdahl, G. (1988). *Palaeoclimatic and Palaeoecological Studies based on Subfossil Insects from the Late Weichselian Sediments in Southern Sweden*. Lundqua Thesis **22**. Lund, Sweden.
- Lindroth, C.H. (1945). *Die Fennoskandischen Carabidae I-II*. Goteborgs K. Vetensk. o VitterhSamh. Handl.(6) B, 4. Goteborg. [Reprinted as English translation , 1992, Intercept, Andover].
- Lindroth, C.H. (1974). *Coleoptera: Carabidae. Handbooks for the Identification of British Insects* **IV**, **2**. Royal Entomological Society of London.
- Lindroth, C.H. (1985). *The Carabidae (Coleoptera) of Fennoscandia and Denmark*. Fauna Entomologica Scandinavica, **15** (1). E.J. Brill, Leiden.
- Lindroth, C.H. (1986). *The Carabidae (Coleoptera) of Fennoscandia and Denmark*. Fauna Entomologica Scandinavica **15** (2). E.J. Brill/Scandinavian Science Press Ltd, Leiden/Copenhagen.
- Brown, J.H. & Lomolino, M.V. (1998). *Biogeography* (2nd ed.). Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts.
- Lucht, W.H. (1987). *Die Käfer Mitteleuropas, Katalog*. Goecke & Evers, Krefeld.
- Luff, M.L. (1998). *Provisional atlas of the ground beetles (Coleoptera, Carabidae) of Britain*. Centre for Ecology & Hydrology, Biological Records Centre, Abbots Ripton.

- Luff, M.L., Eyre, M.D. & Rushton, S.P. (1989). Classification and Ordination of Habitats of Ground Beetles (Coleoptera, Carabidae) in North-East England. *Journal of Biogeography* **16**, 121-130.
- Luff, M.L., Eyre, M.D. & Rushton, S.P. (1992). Classification and prediction of grassland habitats using ground beetles (Coleoptera, Carabidae). *Journal of Environmental Management* **35**, 301-315.
- Machatschke, J.M. (1969). Scarabaeidae. In H. Freude, K.W. Harde & G.A. Lohse (eds.), *Die Käfer Mitteleuropas* **8**. Goecke & Evers, Krefeld, 266-367.
- Majerus, M.E.N., Roy, H.E. & Mabbott, P. (2006). The harlequin ladybird, *Harmonia axyridis* (Pallas) (Col., Coccinellidae) arrives in Britain. *Entomologist's Monthly Magazine* **42**, 87-92.
- Mani, M.S. (1968). *Ecology and Biogeography of High Altitude Insects*. W. Junk, The Hague.
- Manly, B. (1997). *Randomization, Bootstrap, and Monte Carlo Methods in Biology* (2nd ed.). Chapman & Hall, London.
- Marra, M.J., Smith, E.G.C., Shulmeister, J., & Leschenc, R. (2004). Late Quaternary climate change in the Awatere Valley, South Island, New Zealand using a sine model with a maximum likelihood envelope on fossil beetle data. *Quaternary Science Reviews* **23**, 1637-1650.
- Mittelbach, G.G., Schemske, D.W., Cornell, H.V., Allen, A.P., Brown, J.M., Bush, M.B., Harrison, S.P., Hurlbert, A.H., Knowlton, N., Lessios, H.A., McCain, C.M., McCune, A.R., McDade, L.A., McPeck, M.A., Near, T.J., Price, T.D., Ricklefs, R.E., Roy, K., Sax, D.F., Schluter, D., Sobel, J.M. & Turelli, M. (2007). Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters* **10** (4), 315–331.
- Mjöberg, E. (1915). Über die Insektenreste der sogenannten "Härnögyttja" im nordlichen Schweden. *Sveriges geologiska undersökning. Afhandlingar och Uppsatser* **C268**, 1-14.
- Morgan, A. (1970). *Weichselian insect faunas of the English Midlands*. Unpublished PhD thesis, University of Birmingham, UK.
- Moseley, K.A. (1982). *Climatic changes in the Early Devensian cold stage interpreted from Coleopteran assemblages*. Unpublished PhD thesis, University of Birmingham, UK.
- Nilsson, A.N. (1983). *Ilybius vittiger* (Gyllenhal) n. comb. (Coleoptera: Dytiscidae), with description of its larval instars. *Entomologia scandinavia* **14**, 49-56.
- Nilsson, A.N. & Holmen, M. (1995). *The aquatic Adephaga (Coleoptera) of Fennoscandia and Denmark. II. Dytiscidae*. Fauna Entomologica Scandinavica **32**. E.J. Brill, Leiden.
- Nyman, M., Korhola, A. & Brooks, S.J. (2005). The distribution and diversity of Chironomidae (Insecta: Diptera) in western Finnish Lapland, with special emphasis on shallow lakes. *Global Ecology and Biogeography* **14**, 137-153.
- Odgaard, B.V. (1999). Fossil pollen as a record of past biodiversity. *Journal of Biogeography* **26**, 7-17.
- Osborne, P.J. (1969). An insect fauna of Late Bronze Age date from Wilsford, Wiltshire. *Journal of Animal Ecology*, **38**, 555-566.
- Osborne, P.J. (1972). Insect faunas of Late Devensian and Flandrian age from Church Stretton, Shropshire. *Philosophical Transactions of the Royal Society of London* **B263**, 327-367.
- Osborne, P.J. (1980). The Late Devensian-Flandrian transition depicted by serial insect faunas from West Bromwich, Staffordshire, England. *Boreas* **9**, 139-147.
- Osborne, P.J. (1989). Insects. In P. Ashbee, M. Bell & E. Proudfoot (eds.), *Wilsford Shaft. Excavations 1960-62*. English Heritage, London, 96-99 + fiche C1-7.
- Parmesan, C., Gaines, S., Gonzalez, L., Kaufman, D.M., Kingsolver, J., Peterson, A.T. & Sagarin, R. (2005). Empirical perspectives on species borders: from traditional biogeography to global change. *Oikos* **108**, 58-75.
- Perry, D.W. (1986). *The Analysis of sub-fossil insect assemblages: a Numerical Approach*. Unpublished PhD thesis, University of Birmingham, UK.

- Perry, D.W., Buckland, P.C. & Snæsdóttir, M. (1985). The Application of Numerical Techniques to Insect Assemblages from the Site of Stóraborg, Iceland. *Journal of Archaeological Science*, **12**, 335-345.
- Ponel, P. (1995). Rissian, Eemian and Würmian Coleoptera assemblages from La Grande Pile (Vosges, France). *Palaeogeography, Palaeoclimatology, Palaeoecology* **114**, 1-41.
- Porch, N.A. (2006). *A method of reconstructing the Quaternary climates of Australia using fossil beetles*. Unpublished PhD thesis, Geography and Environmental Science. Monash University, Victoria, Australia.
- Prentice, I.C. (1985). Pollen representation, source area, and basin size: toward a unified theory of pollen analysis. *Quaternary Research* **23**, 76-86.
- Pross, J. & Klotz, S. (2002). Paleotemperature calculations from the Prae-tiglian/Tiglian (Plio-Pleistocene) pollen sequence of Lieth, northern Germany: Implications for the climate evolution of northwestern Europe. *Global and Planetary Change* **34**, 253-267.
- Robinson, M.A. (2001). Insects as palaeoenvironmental indicators. In D.R. Brothwell & A.M. Pollard (eds.), *Handbook of archaeological sciences*. J. Wiley & sons, Chichester, 121-133.
- Rosén, P., Segerström, U., Eriksson, L., Renberg, I. & Birks, H.J.B. (2001). Holocene climate change reconstructed from diatoms, chironomids, pollen and near-infrared spectroscopy at an alpine lake (Sjuodjijaure) in northern Sweden. *The Holocene* **11**, 551-562.
- Rosenzweig, M.L. (1995). *Species diversity in space and time*. Cambridge University Press.
- Sadler, J.P. (1987). *The analysis of insect remains from Norse sites in the former Western Settlement of Greenland*. Unpublished MSc thesis, University of Birmingham, UK.
- Sadler, J.P. (1991). *Archaeological and palaeoecological implications of palaeoentomological studies in Orkney and Iceland*. Unpublished PhD thesis, University of Sheffield, UK.
- Sadler, J.P., Buckland, P.C. & Rains, M. (1992). BUGS: an entomological database. *Antenna* **16**, 158-166.
- Schiffer, M.B. (1987). *Site Formation Processes of the Archaeological Record*. University of New Mexico Press, Albuquerque.
- Shotton, F.W., Osborne, P.J. & Grelg, J.R.A. (1977). The fossil content of a Flandrian deposit at Alcester. *Proceedings of the Coventry and District Natural History and Scientific Society* **5**, 19-32.
- Simmons, I.G. (1993). *Interpreting nature : cultural constructions of the environment*. Routledge, London.
- Sinka, K.J. & Atkinson, T.C. (1999). A mutual climatic range method for reconstructing palaeoclimate from plant remains. *Journal of the Geological Society of London* **156**, 381-396.
- Skidmore, P. (1991). *Insects of the British Cow-Dung Community*. Field Studies Council, Occasional Publications **21**.
- Skidmore, P. (unpubl.). *Keys to the British Beetle (Coleoptera) Fauna*. (Included in the BugsCEP database).
- Smith, D.N., Gaffney, V., Grossman, D., Howard, A.J., Milosevic, A., Ostir, K., Todobnikar, T., Smith, W., Tetlow, E., Tingle, M. & Tinsley, H. (2006). Assessing the later prehistoric environmental archaeology and landscape of the Cetina valley, Croatia. *Environmental Archaeology* **11**, 171-187.
- Southwood, T.R.E. (1978). *Ecological methods, with particular reference to the study of insect populations* (2nd ed.). John Wiley & Sons, New York.
- Southwood, T.R.E. & P.A. Henderson (2000). *Ecological methods* (3rd ed.). Blackwell Science, Malden, Massachusetts, USA.
- Stafford, F. (1971). Insects of a medieval burial. *Science and Archaeology* **7**, 6-10.
- Stern, N. (2006). *The Stern Review Report on the Economics of Climate Change*. UK Treasury, London.

- Stewart, J.R. (2003). Comment on “Buffered Tree Population Changes in a Quaternary Refugium: Evolutionary Implications”. *Science* **299**, 825.
- Sugita, S., Gaillard, M.J. & Broström, A. (1999). Landscape openness and pollen records: a simulation approach. *The Holocene* **9**, 409-421.
- Sundborg, P. (2006). *Ska postmodernismen ta över kulturarvssektorn?* Debatt artikel i DIK Forum **13**.
- Tegby, I. (2004). *Biostratigrafisk undersökning av en myr i Hemavan*. CD-dissertation, Archaeology, Umeå University, Sweden.
- Thomas, J.A., Telfer, M.G., Roy, D.B., Preston, C.D., Greenwood, J.J.D., Asher, J., Fox, R., Clarke, R.T., & Lawton, J.H. (2004). Comparative Losses of British Butterflies, Birds, and Plants and the Global Extinction Crisis. *Science* **303**, 1879-1881.
- Trigger, B. (1989). *A history of archaeological thought*. Cambridge University Press, Cambridge.
- Tzedakis, P.C., Lawson, I.T., Frogley, M.R., Hewitt, G.M. & Preece, R.C. (2002). Buffered Tree Population Changes in a Quaternary Refugium: Evolutionary Implications. *Science* **297**, 2044-2047.
- Tzedakis, P.C., Lawson, I.T., Frogley, M.R., Hewitt, G.M. & Preece, R.C. (2003). Response to Comment on “Buffered Tree Population Changes in a Quaternary Refugium: Evolutionary Implications”. *Science* **299**, 825b.
- Vera, F.W.M. (2000). *Grazing ecology and forest history*. CABI, Wallingford.
- Vetenskapsrådet (2005). Om forskningens infrastrukturer inom humaniora och samhällsvetenskap i Sverige. Report Dnr 131-2004-8236.
- Vickers, K.A. (2006). *The Palaeoentomology of the North Atlantic Islands*. PhD thesis, Archaeology and Prehistory, University of Sheffield, UK.
- Vickers, K.A. & Buckland, P.I. (*in prep.*). Predicting island beetle faunas by their climate ranges - tabula-rasa/refugia theory in the North Atlantic.
- Wagner, P.E. (1997). Human Impact or Cooling Climate? The "Little Ice Age" and the beetle fauna of the British Isles. In A.C. Ashworth, P.C. Buckland & J.P. Sadler (eds.), *Studies in Quaternary Entomology - An Inordinate Fondness for Insects*. *Quaternary Proceedings* **5**, 269-276.
- Webb, J.R. & Lott, D.A. (2006). The development of ISIS: a habitat-based invertebrate assemblage classification system for assessing conservation interest in England. *Journal of Insect Conservation* **10**, 179-188.
- Williams, J.W., Shuman, B.N. & Webb III, T. (2001). Dissimilarity analyses of late-quaternary vegetation and climate in eastern North America. *Ecology* **82**, 3346-3362.
- Witte, H.J.L., Coope, G.R., Lemdahl, G. & Lowe, J.J. (1998). Regression coefficients of thermal gradients in northwestern Europe during the last glacial - Holocene transition using beetle MCR data. *Journal of Quaternary Science* **13**, 435-446.
- Wolfe, A.P. (2003). Diatom community responses to late Holocene climatic variability, Baffin Island, Canada: a comparison of numerical methods. *The Holocene* **13**, 29-37.
- Yoffee, N. & Sherratt, A (1993). *Archaeological theory: who sets the agenda?* Cambridge University Press, Cambridge.
- Yuan Zhuo, D. (Unpubl.) Masters Thesis, Dept. Computer Science, University of Sheffield, UK.

Online information sources

- ADS – Archaeological Data Service, Arts and Humanities Data Service, York. <http://ads.ahds.ac.uk>
- ArtPortalen (2006). The Species Gateway, Swedish Species Information Centre and Swedish Environmental Protection Agency. <http://www.artportalen.se/>
- Chi, P. & Russell, C.T. (1999). Statistical Methods for Data Analysis in Space Physics [Online content]. <http://www-ssc.igpp.ucla.edu/personnel/russell/ESS265/Ch9/autoreg/autoreg.html> [Accessed 2007-02-15].
- EUNIS (2006) <http://eunis.finsiel.ro/eunis/> European Environment Agency.

- Fauna Europaea Web Service (2004). Fauna Europaea version 1.1. <http://www.faunaeur.org>
- Finnish Meteorological Institute (FMI). http://www.fmi.fi/weather/climate_6.html [Accessed 2006-12-08].
- GBIF (2006). Global Biodiversity Information Facility. <http://www.gbif.org/>
- Ground Beetles of Ireland (2006). National Museums, Northern Ireland. <http://www.habitas.org.uk/groundbeetles/>
- The Harlequin Ladybird Survey (and associated reports). <http://www.harlequin-survey.org/>
- Microsoft Knowledge Base article 242933. <http://support.microsoft.com/kb/242933/en-us>
- SMHI (2005). Sveriges meteorologiska och hydrologiska institut, downloadable and online data. <http://www.smhi.se/> [Accessed 2007-02-14].
- Wikipedia, Holocene extinction event. http://en.wikipedia.org/wiki/Holocene_extinction_event. [Accessed 2007-03-12].
- Wikipedia, Relational databases. http://en.wikipedia.org/wiki/Relational_database. [Accessed 2007-02-14].

Internet sites mentioned

- Bugs Coleopteran Ecology Package Website. <http://www.bugscep.com>
- Genesys, University of Sheffield, UK. <http://www.genesys.shef.ac.uk/>
- GEON, The Geosciences Network. <http://www.geongrid.org/>
- Google Maps. <http://maps.google.com/maps>
- Microsoft Developers Network, SQL-Server: <http://msdn2.microsoft.com/en-us/sql>
- MySQL: <http://www.mysql.com>
- PostgreSQL: <http://www.postgresql.org>
- World Data Center for Paleoclimatology (NOAA/NCDC). <http://www.ncdc.noaa.gov/paleo/paleo.html>

Software and databases

- Bronk Ramsey, C. (1995). Radiocarbon calibration and analysis of stratigraphy: The OxCal program. *Radiocarbon* **37** (2), 425-430
- Hill, M.O. (1979). TWINSpan – a FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. *Ecology and Systematics*, Cornell University, Ithaca, New York.
- Hood, G.M. (2003) PopTools version 2.5.9. Available on the internet. URL <http://www.cse.csiro.au/poptools>
- Köhler, F. & Klausnitzer, B. (eds) (1998). Verzeichnis der Käfer Deutschlands (Datenbanken) [The Coleoptera of Germany, database]. <http://www.koleopterologie.de/verzeichnis-der-kaefer-deutschlands/index.html>
- MapMate (2000). ©Teknica Ltd. <http://www.mapmate.co.uk/>
- Recorder 6 (2006). Built to NBN standards by Dorset Software Ltd., ©JNCC, 2006. <http://www.nbn.org.uk/recorder>

BugsCEP version used for calculations in this thesis

- Buckland P.I. & Buckland P.C. (2006). *Bugs Coleopteran Ecology Package* (Versions: BugsCEP v7.62; Bugsdata v7.10; BugsMCR v2.02; BugStats v1.22).

Archaeology and Environment

Published by the Department of Archaeology and Sámi Studies, Umeå University

Series editor: Roger Engelmark & Thomas Larsson

1. Ramqvist, Per H: Gene. On the origin, function and development of sedantary Iron Age settlement in Northern Sweden. 1983.
2. Baudou, Evert. (ed): Papers in Northern Archaeology. 1984.
3. Zachrisson, Inger: De samiska metalldepåerna år 1000-1350 i ljuset av fyndet från Mörtrträsket, Lappland (The Saami Metal Deposits A.D.1000-1350 in the light of the find from Mörtrträsket, Lapland). 1984.
4. Backe, Margareta. et al. (eds) In Honorem Evert Baudou. 1985.
5. Forsberg, Lars: Site Variability and Settlement Patterns. An analysis of the hunter-gatherer settlement system in the Lule River Valley, 1500 B.C.- B.C./A.D. 1985.
6. Larsson, Thomas B: The Bronze Age Metalwork in Southern Sweden. Aspects of social and spatial organization 1800-500 B.C. 1986.
7. Segerström, Ulf: The Post-Glacial History of Vegetation and Agriculture in the Luleälv River Valley. 1990.
8. Hulthén, Birgitta: On Ceramic Ware in Northern Scandinavia During the Neolithic Bronze and Early Iron Age. A Ceramic-Ecological Study. 1991.
9. Nockert, Margareta: Högom. Part II. The Högom find and other Migration Period textiles and costumes in Scandinavia. 1991.
10. Aronsson, Kjell-Åke: Forest Reindeer Herding A.D. 1-1800. An archaeological and palaeoecological study in Northern Sweden. 1991.
11. Baudou, Evert (ed.): Nordic TAG. Report from the 2nd Nordic TAG conference, Umeå 1987. 1991.
12. Holm, Lena: The Use of Stone and Hunting of Reindeer. A study of stone tool manufacture and hunting of large mammals in the central Scandes c. 6000 - 1 BC. 1992.
13. Ramqvist, Per H: Högom. Part I. The excavations 1949-1984. 1992.
14. Viklund, Karin. Cerals, Weeds and Crop Processing in Iron Age Sweden. 1998.
15. Viklund, Karin (ed.). Nordic Archaeobotany - NAG 2000. 2002.
16. Olofsson, Anders. Pioneer Settlement in the Mesolithic of Northern Sweden. 2003.
17. Hulthén, Birgitta & Larsson, Thomas B: Vistad '88 Revisited. Ceramological Analyses and Lusatian Connections. 2004.
18. Loeffler, David: Contested landscapes/Contested Heritage. History and heritage in Sweden and their archaeological implications concerning the interpretation of the Norrlandian past. 2005.
19. Engelmark, Roger, Larsson, Thomas B & Rahtje, Lillian (redaktörer). en lång historia...festskrift till Evert Baudou på 80-årsdagen. 2005.
20. Rogersdotter, Elke. The Forgotten: an Approach on Harappan Toy Artefacts. 2006.
21. Roger Engelmark and Johan Linderholm (eds.). Proceedings from the VIII Nordic conference on the applications of Scientific Methods in Archaeology Umeå 2001. 2006.
22. Mulk, Inga-Maria & Bayliss-Smith, Tim. Rock Art and Sami Sacred Geography in Badjelánda, Laponia, Sweden. 2006.



ISBN 978-91-7264-298-0
ISSN 0281-5877