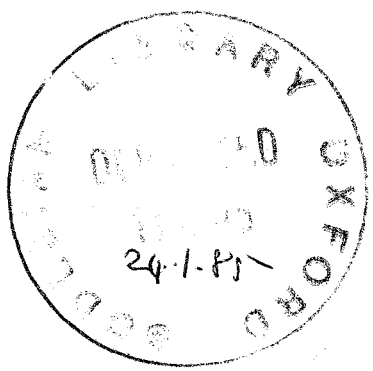
THE ANALYSIS OF QUATERNARY CAVE SEDIMENTS AND ITS BEARING UPON PALAEOLITHIC ARCHAEOLOGY, WITH SPECIAL REFERENCE TO SELECTED SITES FROM WESTERN BRITAIN

S.N. Collcutt

St. John's College



Volume 1

Thesis submitted for the Degree of

Doctor of Philosophy

Oxford University

1984

I have long discovered that geologists never read each other's works, and that the only object in writing a book is a proof of earnestness, and that you do not form your opinions without undergoing labour of some kind.

Charles Darwin (letter to J.M. Herbert, c.1843)

In the hope of proving Darwin wrong ...

CONTENTS

Volume 1

CONTENTS	iii
LIST OF FIGURES	ix
ACKNOWLEDGEMENTS	xii
ABSTRACT	xiv
1. INTRODUCTION	1
1.1. Caves and the Development of Quaternary Studies	1
1.2. The Development of Cave Sedimentology	10
1.3. The Present Thesis	15
1.4. A Basic Framework	22
PART I - THEORY	
2. THE CAVE ENVIRONMENT	41
2.1. General	41
2.2. Cave Meteorology	43
2.3. Zonation	80
3. THE PROPERTIES OF SEDIMENTS	87
3.1. Particles and Sediments	87
3.2. Patterns and Facies	102
3.3. A Case Study of Sediment Patterning: Three Cantabrian Caves	112
4. FIELDWORK	127
5. BOUNDARY FEATURES AND STRATIGRAPHY	134
5.1. Space and Time	134
5.2. The International Stratigraphic Guide	145

6. QUANTIFICATION 168
6.1. Why Quantify? 168
6.2. Sampling Theory 173
6.3. Numbers, Scales and Indices 196

7. ANALYTICAL TECHNIQUES 229
7.1. Laboratory Analysis 229
7.2. A Case Study: ΔpH and Clay Mineral Alteration 232

8. THE RECOGNITION OF THE EFFECTS OF ICE 253
8.1. Ice in the European Quaternary 253
8.2. Ice in Sediments 257
8.3. Rock Shattering 281

9. MASS MOVEMENT 312
9.1. Generalities 312
9.1.1. (Statement) 312
9.1.2. Definition 312
9.1.3. Classification 313
9.2. Mass Movement with Little Lateral Displacement 315
9.2.1. (General) 315
9.2.2. The Collapse of Hard Rock 316
9.2.3. The Collapse of Indurated Sediment 324
9.2.4. The Collapse of Unconsolidated Sediment 325
9.3. Slow Mass Movement 326
9.3.1. General 326
9.3.2. The Concepts of Solifluction and Gelifluction 329
9.3.3. Sedimentary Features Resulting from Slow Mass Movement 334
9.3.4. Mass Movement Involving Very Fine Grained Sediments 365
9.4. Fast Mass Movement 371
9.4.1. General 371
9.4.2. Debris Flow 371
9.4.3. Avalanche 397

10. MECHANICAL PROCESSES ASSOCIATED WITH WATER MOVEMENT 399
10.1. Wash Processes 399
10.2. Permanent Freshwater Bodies 418
10.3. Sub-surface Processes 439
10.4. Marine Processes 450

Volume 2

11. WIND ACTION 457
11.1. General Aeolian Processes 457
11.2. The Environmental Implications of Aeolian Deposits 462
11.3. Criteria for the Recognition of Wind Action 473

12. ALTERATION AND SOILS 491
12.1. Processes Involved in Alteration 491
12.2. The Recognition of Alteration Involving Carbonates 522
12.3. Alteration Systems and Soils 530

13. SPELEOTHEMS 552

14. BIOLOGICAL MATERIAL 565
14.1. Is Biological Material from Caves Worth Collecting? 565
14.2. The Recovery of Biological Material from Caves 568
14.3. Taphonomy and Palaeoecology 573
14.4. Cave Palynology 584

15. PROVENANCE AND PALAEOGEOGRAPHY 594

16. DIAGENESIS 602

17. CORRELATION 612

18. DATING METHODS 634

19. SYNTHESIS OF RESULTS 645
19.1. Multidisciplinary Research 645
19.2. A Case Study: The Multidisciplinary Study of a Cave Site - L'Hortus (Hérault) 645
19.3. Professionalism 657

20. ARCHAEOLOGICAL CONCEPTS 661

PART II - STUDY SITES

21. SOUTH DEVON 688

21.1. Introduction 688

21.2. Tornewton Cave 691

21.2.1. Tornewton Cave - Situation and Morphology 691

21.2.2. Tornewton Cave - Excavations 695

21.2.3. Tornewton Cave - Surviving Interior Deposits 709

21.2.4. Tornewton Cave - Surviving Deposits in the Middle Tunnel 716

21.2.5. Tornewton Cave - Surviving Exterior Deposits 719

21.2.6. Tornewton Cave - Discussion 721

21.3. Three Holes Cave 739

21.3.1. Three Holes Cave - Morphology 739

21.3.2. Three Holes Cave - Previous Excavations 742

21.3.3. Three Holes Cave - Recent Observations 750

21.3.4. Three Holes Cave - Interpretation of the Deposits 764

21.3.5. Three Holes Cave - Discussion 775

21.4. The Torbryan Valley - Conclusion 787

21.5. Joint Mitnor Cave 789

21.5.1. Joint Mitnor Cave - Situation and Morphology 789

21.5.2. Joint Mitnor Cave - Excavations and Previous Study 792

21.5.3. Joint Mitnor Cave - Recent Observations 795

21.5.4. Joint Mitnor Cave - Discussion 805

21.6. The Chudleigh Caves 818

21.6.1. Chudleigh Rocks 818

21.6.2. Pixie's Hole - Morphology 821

21.6.3. Pixie's Hole - Previous Examinations 824

21.6.4. Pixie's Hole - Recent Excavations 828

21.6.5. Pixie's Hole - Interpretation of the Deposits 842

21.6.6. Pixie's Hole - Archaeological Considerations 853

21.6.7. Other Chudleigh Sites 886

Volume 3

22. MENDIP 894

22.1. The Mendip Region 894

22.2. Sun Hole 898

22.2.1. Sun Hole - Introduction 898

22.2.2. Sun Hole - Lithostratigraphy 900

22.2.3. Sun Hole - Sedimentary Processes 914
22.2.4. Sun Hole - The Pleistocene Fauna 924
22.2.5. Sun Hole - Archaeology and Radiocarbon Dating 928
22.2.6. Sun Hole - Discussion (First Approximation) 930
22.2.7. Sun Hole - Discussion (Second Approximation) 934
22.3. The Caves of Wookey Hole Ravine 948
22.3.1. Wookey Hole Ravine - Introduction 948
22.3.2. Badger Hole 951
22.3.3. Rhinoceros Hole 966
22.3.4. Hyaena Den 976
22.3.5. The Caves of Wookey Hole Ravine - Discussion 980
22.4. Westbury-sub-Mendip Fissure 989
22.4.1. Westbury-sub-Mendip Fissure - Introduction 989
22.4.2. Westbury-sub-Mendip Fissure - General Description of the Deposits 992
22.4.3. Westbury-sub-Mendip Fissure - Aspects of the East Basin Deposits 1000
22.4.4. Westbury-sub-Mendip Fissure - Aspects of the West Basin Deposits 1010
22.4.5. Westbury-sub-Mendip Fissure - Postscript 1024
22.5. Mendip - Conclusion 1025

23. GOWER 1027
23.1. Introduction 1027
23.2. Bacon Hole 1029
23.2.1. Bacon Hole - Morphology 1029
23.2.2. Bacon Hole - Previous Excavations 1033
23.2.3. Bacon Hole - Recent Excavations 1035
23.2.4. Bacon Hole - Interpretation of the Sequence 1062
23.2.5. Bacon Hole - Archaeological Considerations 1089
23.3. Beyond Bacon Hole: A Brief Survey of the Quaternary of the South Gower Coast 1093

24. NORTH WALES 1108
24.1. Introduction 1108
24.2. Pontnewydd Cave - Summary 1118
24.3. Pontnewydd Cave - Lithostratigraphy 1120
24.4. Pontnewydd Cave - Correlation 1151
24.5. Pontnewydd Cave - Lithogenesis: General Considerations 1172
24.5.1. Sediment Sources 1172

24.5.2. Speleothems 1180
24.5.3. Debris Flow 1181
24.5.4. Diagenesis and Carbonate Mobility 1186
24.6. Development of the Pontnewydd Sequence 1190
24.6.1. The Siliceous member 1190
24.6.2. The Intermediate complex 1196
24.6.3. The Lower Breccia bed 1200
24.6.4. The Stalagmite lithozone and Silt beds 1201
24.6.5. The Upper Breccia bed 1203
24.6.6. The Red Cave Earth bed 1204
24.6.7. The Upper Deposits 1205
24.7. Pontnewydd Cave - Discussion 1208

25. SUMMARY AND CONCLUSION 1219

APPENDIX: Simulation of Particle Size Distributions 1231

BIBLIOGRAPHY 1242

SITE INDEX 1299

LIST OF FIGURES

Fig.1	Theoretical landscape with caves	26
Fig.2	Theoretical chimney effect	45
Fig.3	Theoretical drainage effect	48
Fig.4	Theoretical entrainment effect	50
Fig.5	Cave meteorology - temperate case (summer)	64
Fig.6	Cave meteorology - temperate case (transitional)	64
Fig.7	Cave meteorology - temperate case (winter)	64
Fig.8	Cave meteorology - cold climate case (summer)	68
Fig.9	Cave meteorology - cold climate case (transitional)	68
Fig.10	Cave meteorology - cold climate case (winter)	68
Fig.11	Cantabrian caves - residual facies analysis	117-118
Fig.12	Cantabrian caves - upward transition matrix	120
Fig.13	Cantabrian caves - upward transition matrix	122
Fig.14	Cantabrian caves - upward transition matrix	123
Fig.15	Contact geometry	143
Fig.16	Theoretical particle size distribution	219
Fig.17	Delta-pH/colloid plot - Pech de l'Azé II, Section 3	236
Fig.18	Delta-pH/colloid plot - Caminade-Est	237
Fig.19	Delta-pH/colloid plot - Port-Morvan	239
Fig.20	Delta-pH/pH relationships	241
Fig.21	Delta-pH/pH relationships	243
Fig.22	Condition of limestone clasts - Combe-Grenal	298
Fig.23	[Plate] Ebbor scree	338
Fig.24	[Plate] G.B. Cave - exterior collapse	385
Fig.25	[Plate] G.B. Cave - interior debris flow	387
Fig.26	[Plate] G.B. Cave - details of debris flow	388
Fig.27	Tornewton Cave	693
Fig.28	Tornewton Cave - comparison of reported stratigraphies	707
Fig.29	Tornewton Cave - surviving interior deposits	710
Fig.30	Tornewton Cave - deposits of the Middle Tunnel and the exterior	717
Fig.31	Tornewton Cave - particle size distributions	727

Fig.32	Three Holes Cave - plan	741
Fig.33	Three Holes Cave - Widger's accounts	744
Fig.34	Three Holes Cave - planes A and B	752
Fig.35	Three Holes Cave - plane C	759
Fig.36	Three Holes Cave - plane D	763
Fig.37	Three Holes Cave - particle size distributions	777
Fig.38	Three Holes Cave - handaxe cast	782
Fig.39	Three Holes Cave - unstruck Levallois core	785
Fig.40	Joint Mitnor Cave - plan	791
Fig.41	Joint Mitnor Cave	794
Fig.42	Joint Mitnor Cave	796
Fig.43	Pixie's Hole - plan	822
Fig.44	Pixie's Hole - section O/P (looking south)	830
Fig.45	Pixie's Hole - section 4/5 (looking west)	834
Fig.46	Pixie's Hole - section N/M (looking north)	834
Fig.47	Pixie's Hole - section 4/5 (looking west)	839
Fig.48	Pixie's Hole - particle size distributions	850
Fig.49	Pixie's Hole - clast orientation: debris flow context	857
Fig.50	Pixie's Hole - clast orientation contours: debris flow context	858
Fig.51	Pixie's Hole - orientation rose (percentage): debris flow context	860
Fig.52	Pixie's Hole - heavy end orientation rose: debris flow context	861
Fig.53	Pixie's Hole - clast orientation: slope wash context	864
Fig.54	Pixie's Hole - clast orientation contours: slope wash context	865
Fig.55	Pixie's Hole - orientation rose (percentage): slope wash context	866
Fig.56	Pixie's Hole - heavy end orientation rose: slope wash context	867
Fig.57	Pixie's Hole - plan of hearth area	870
Fig.58	Pixie's Hole - vertical plot of artefacts in squares U5 and T5	872
Fig.59	Pixie's Hole - vertical section 4/5 across hearth area	872
Fig.60	Sun Hole - main sections	902
Fig.61	Sun Hole - Unit III sands	904
Fig.62	Sun Hole - Unit II sands	908
Fig.63	Sun Hole - Unit I sands	911
Fig.64	Sun Hole - sedimentary events	915
Fig.65	Sun Hole - particle size distributions	922
Fig.66	Sun Hole - fauna from the 1977-1980 excavations	927
Fig.67	Badger Hole - plan	952
Fig.68	Badger Hole - schematic sections	954
Fig.69	Rhinoceros Hole - southerly sections	968
Fig.70	Wookey Hole area - sedimentary stocks	982

Fig.71	Westbury-sub-Mendip Fissure - schematic section	993
Fig.72	[Plate] Westbury-sub-Mendip - Brighton Beach	1006
Fig.73	[Plate] Westbury-sub-Mendip - Brighton Beach	1007
Fig.74	Westbury-sub-Mendip - West Basin	1011
Fig.75	[Plate] Westbury-sub-Mendip - deformation structures	1015
Fig.76	Bacon Hole - highly schematic plan	1030
Fig.77	Bacon Hole - platform section	1037
Fig.78	Bacon Hole - interior sections	1058
Fig.79	The south Gower coast	1094
Fig.80	Pontnewydd Cave - plan	1122
Fig.81	Pontnewydd Cave - schematic composite lithostratigraphy	1123
Fig.82	Pontnewydd Cave - Deep Sounding and South Fissure	1128
Fig.83	Pontnewydd Cave - Area B/C	1137
Fig.84	Pontnewydd Cave - coded sediment characteristics	1153
Fig.85	Pontnewydd Cave - ranked non-carbonate sand grades	1155
Fig.86	Pontnewydd Cave - lithologic correlation: similarity coefficients	1157
Fig.87	Pontnewydd Cave - lithologic correlation: rank correlation	1157
Fig.88	Pontnewydd Cave - lithologic correlation: similarity coefficients	1159
Fig.89	Pontnewydd Cave - lithologic correlation: rank correlation	1159
Fig.90	Pontnewydd Cave - lithologic correlation: similarity coefficients	1160
Fig.91	Pontnewydd Cave - lithologic correlation: rank correlation	1160
Fig.92	Pontnewydd Cave - lithostratigraphic classification	1163
Fig.93	Pontnewydd Cave - schematic lithostratigraphy	1164
Fig.94	[Plate] Pontnewydd Cave - representative limestone clasts and pebbles	1173
Fig.95	Pontnewydd Cave - organic matter	1174
Fig.96	Pontnewydd Cave - carbonate content	1175
Fig.97	Pontnewydd Cave - size distribution of non-carbonate sands	1176
Fig.98	Pontnewydd Cave - coarser particles	1177
Fig.99	Pontnewydd Cave - particle size distributions	1178
Fig.100	Pontnewydd Cave - pebble fracturing	1179
Fig.101	[Plate] Pontnewydd Cave - experimental debris flow	1185

ACKNOWLEDGEMENTS

A Major State Scholarship (Department of Education and Science) enabled me to begin my research at Oxford. Grants from the Principal's Fund (St. John's College, Oxford) and the Meyerstein Fund (Oxford University) helped towards the cost of excavation at Pixie's Hole. A Visiting Fellowship from the Belgian Ministry of Culture, sponsored by Dr. M. Dewez, gave me the opportunity to visit cave sites in the region around Liège. My work at Bacon Hole and Westbury-sub-Mendip was funded by the British Museum (Natural History), and that at Pontnewydd was partially funded by the National Museum of Wales. A Grant in Aid from the Leakey Foundation, in general support of my research into cave sediments, provided a most solid basis for my work. To all these organisations and their administrators, I offer my most sincere thanks.

Examination of the Torbryan sites was undertaken with the kind permission of the owners, the Starke Family of Tornewton House. Mrs. A. Longman provided both advice and much needed transport during this work. I am most grateful.

The William Pengelly Cave Research Association, and in particular Dr. D. Curry and Mr. W. Joint, made it possible for me to work in Joint Mitnor Cave; in addition, I was provided with accommodation, access to museum collections and storage space for my sediment samples. I extend my thanks, and congratulations, to this most remarkable Association.

Lord Clifford of Chudleigh very kindly permitted me to dig in the caves in Chudleigh Rocks. The Boulton Family of Chudleigh Nursery showed us a degree of hospitality which might be described as unreserved 'adoption'. I am truly grateful for our visits to Chudleigh which probably constituted the most pleasant, and by no means the least educational, part of my fieldwork.

Mr. D. Thackray of the National Trust has shown a most welcome interest in our work at Sun Hole. Dr. J. Rodgers, R. Harrison, D. Wassell and N. Hawkes have all helped with the excavation of this uncompromising site. To all these hardy people, I extend my sincere thanks.

I am obliged to Wookey Hole Caves Ltd. for permission to examine the sites on their land and for access to their museum collection.

During my research, I have been lucky enough to have worked with a large number of quaternary specialists, many of whom are acknowledged in the pages of this thesis. I am deeply grateful for the advice and aid which they have offered and I am sure that my research has greatly benefited.

A number of individuals have been especially helpful over the years. Dr. A.J. Sutcliffe and the late Professor E.K. Tratman introduced me, with contagious enthusiasm, to their respective 'haunts' of South Devon and Mendip, and afforded me every support in my work. C.J. Hawkes and Dr. W. Stanton will never tire of teaching me about Mendip. Chris Stringer and Andy Currant have shown me that true collaboration is indeed possible, even when one of the parties is as argumentative as myself. My brother, Mike Collcutt, has consistently plugged the holes in my mathematical education. Nick Barton,

whose excellent photographs appear in this thesis, has been in all things a most stalwart friend and colleague. John Dumont processed the photographs in this thesis, Lotti Frascoli gave invaluable advice on translations from German, and all the other 'inmates' of the Donald Baden-Powell Quaternary Research Centre helped to realise this work in one way or another.

I owe a special debt to my supervisor, Dr. Derek Roe; his patient steering is very much more appreciated than I have perhaps led him to believe. Those parts of this thesis which contrive a measure of organisation and lucidity are largely due to his skill and perserverance, whilst those which do not are certainly due to my obstinacy.

Through eight long and sometimes difficult years my wife, Katy, and latterly my daughter, Marianne, have all too often been left to their own devices. And yet, they were always there when I needed them. They have my love and gratitude.

S.N.C.

Oxford, June 1984.

ABSTRACT

The Analysis of Quaternary Cave Sediments and Its Bearing Upon Palaeolithic Archaeology, with Special Reference to Selected Sites from Western Britain.

S.N. Collcutt
Doctor of Philosophy
St. John's College, Oxford
1984

The present thesis constitutes a reassessment of quaternary cave sediment analysis and its relationships to palaeolithic archaeology. Traditionally, archaeological cave sediments have been interpreted in climatic terms, the arguments usually being based upon compositional and textural parameters measured in the laboratory using a fixed batch of analytical techniques. The author demonstrates that this approach is unsound, both because of the simplistic interpretative framework employed and because, even when highly sophisticated techniques are used, it is often difficult to isolate climatic controls, factors which are only a small part of the lithogenetic process. The author has concentrated upon contextual matters, giving priority to the deduction of sedimentary environments through actualistic principles. Great importance is attached to structure, geometry and facies, features which are best observed on site. Thus, the author addresses the primary problem of how a sediment formed. This problem is fundamental to archaeology and all other object-oriented disciplines since, without a solution, one cannot assess the significance of assemblages with any degree of confidence. Furthermore, any climatic information which is indeed available can only be extracted once a firm contextual framework has been established.

The thesis has been split into two Parts. The first Part contains a review of sedimentological theory as applicable to cave deposits. This involves discussion of basic principles (the cave environment, properties of sediments, fieldwork, boundary features, stratigraphy, quantification, analytical techniques), process groupings (mechanisms and effects of ice formation, mass movement, water movement, wind, alteration, chemical precipitation) and more synthetic topics (biological material, provenance, palaeogeography, diagenesis, correlation, dating methods, synthesis of results, archaeological concepts). The second Part comprises a discussion of various British cave sites (South Devon, Mendip, Gower, North Wales), each section being presented, not as a site report, but as an illustration of some of the theoretical points raised in the first Part.

1. INTRODUCTION

1.1. Caves and the Development of Quaternary Studies

. . . . . .
Where Alph, the sacred river, ran
Through caverns measureless to man
Down to a sunless sea.

Samuel Taylor Coleridge, "Kubla Khan"

Man's general attitude towards caves is well illustrated in the above excerpt. The theme of mystery, even sanctity, runs through the literature of both the Western and Eastern Worlds. The theme may even be much older than the written word. Leroi-Gourhan, discussing palaeolithic cave art, argues that decorated caves should be seen as "sanctuaries" and suggests that "the cave as a whole does seem to have had a female symbolic character"(1968:174). In Europe, apart from rare survivals such as Lourdes, the sanctity of caves has been eroded, leaving only folktale or pure superstition. It is remarkable how many caves with local names are referred to as the abode of fairies, mythical heroes or peculiar old ladies. However we may regard caves, it seems difficult for us to ignore them. Some caves even have long documented histories. For instance, the Great Cave of Wookey Hole (Mendip) was granted to the church at Wells in a pre-Doomsday Book (1065) charter and ever since has been frequently referenced (cf. Balch 1929).

From the viewpoint of quaternary studies, it would appear

that man's intuitive appreciation of caves is not too distant from modern scientific concepts. In the past, caves have been seen as permanent doorways, on the threshold between the ordinary and the mysterious, between this world and the next. Now, it is useful to consider caves as metastable 'containers', on the threshold between the energetic, shifting landscape and the subterranean world where change is measured on much longer time-scales.

The transition from folklore to scientific cave research was, as with most other disciplines, a slow process. The tendency for caves to contain fossil bone is perhaps their most obvious attribute. Bones were occasionally recovered during casual visits to caves or, in much greater numbers, during mining operations. Permission to mine saltpetre from the cave of Mas d'Azil (French Pyrenees) was given by Royal Act in 1585. The Eocene-Oligocene fissure deposits of the Quercy were mined for their phosphorites by the seventeenth century, and the 'ochre pits' of the Banwell Caves (Mendip) by the eighteenth. The bones discovered were usually attributed to fabulous creatures, such as dragons or giants (cf. the poem by Grebner quoted in Kurten 1969). Dr. C. Gesner, in his Historiae Animalium (1603), has a whole chapter on cave fossils entitled "Monoceras", the Unicorn, a most popular designation at this time. By the end of the eighteenth century, enough fossil material had been collected to form the basis of less fanciful classification. Linnaeus had already prepared the way with his work on extant animals in Systema naturae. Cuvier's Recherches sur les ossements fossiles de Quadrupèdes was published in 1812 and contained a great majority of cave specimens.

The contribution of cave sites to the palaeontology of

quaternary mammals is commemorated in the names given to these creatures, even though many of these names are now considered invalid. Some animals occurred so frequently in such sites that the word 'cave' was used as the specific name. The den carnivores Crocuta spelaea Goldfuss, Felis spelaea Goldfuss, Ursus spelaeus Rosenmüller & Heinroth and Canis spelaeus Goldfuss are not particularly surprising, but Owen's Lagomys spelaeus and Strongyloceros spelaeus, a steppe pika and a giant deer respectively, show how the type of find site was considered more important than the creature's probable habitat. Sometimes the name of the cave or cave locality has provided the specific or subspecific name: Canis lupus lunellensis Bonifay and Felis lunellensis Bonifay (Lunel-Viel, Hérault); Hyaena makapani Toerien (Makapan, Transvaal); Palaeoloxodon antiquus melitensis Falconer and Leithia (Myoxus) melitensis Leith Adams (Adam's Cave and Maghlak Cave respectively, Malta); Myotragus balearicus Bate (Cala Morlanda, Majorca); Felis gombaszögensis Kretzoi (Gombasek Fissures, Czechoslovakia). Recently, Chaline (1972) has subdivided many of the French pleistocene rodent groups into new subspecies, nearly all of which bear the name of a cave or a cave region. Sometimes homage has been paid to fellow or past researchers: Redunca darti Wells & Cooke (R.A.Dart; holotype - Makapan, Transvaal); Makapania broomi Wells & Cooke (R.Broom; holotype - Makapan, Transvaal); Cervus bucklandi Owen (W.Buckland; holotype - Kirkdale Cave, Yorkshire); Nototherium mitchelli Owen (T.L.Mitchell; holotype - Wellington Valley Caves, Western Australia); Hypnomys waldreni Reumer (W.Waldren; holotype - the fissures of Cap Farrutx, Majorca). The holotypes of many more pleistocene mammals were recovered from caves all over the world. In fact, such finds are not restricted to the

Pleistocene. For example, in 1903 Boyd Dawkins reported on a pliocene fauna from Doveholes (Derbyshire) and the Morganucodon and Kuehneotherium from rhaeto-liassic fissure deposits in South Wales constitute two of the earliest mammals from anywhere in the world (Kermak et al. 1973). Even faunal stages have been defined using cave assemblages, for example, the Villanyian (upper Villafranchian) of Hungary which takes its name from the Villány Fissure.

Quite early in the history of cave exploration it became apparent that, even when all the skeletons of 'fairies' and 'giants' had been allocated to their correct animal taxa, there still remained a few occurrences of unequivocally human bones. At first, these observations could not be reconciled with the philosophical framework then current. The moral dilemma is clearly discernible in J.F.Esper's discussion of his work in the Gailenreuth Cave: "I dare not presume without any sufficient reason these human members to be of the same age as the other animal petrifactions."(1774; translation in Daniel 1950;25). Similar qualms prevented J.MacEnery's work at Kent's Cavern (Devon) from being published until after his death. But other researchers were not so reticent: Tournal, de Serres, de Christol, Dumas, Vatar de Jouannet, de Saussure and Pitore were all prepared to state their belief in the contemporaneity of man, artefacts and extinct animals in the French caves they had been excavating in the early 1800s. P.C.Schmerling was unequivocal in his opinion concerning his finds from Engis (Belgium): "There can be no doubt that the human bones were buried at the same time and by the same cause as the other extinct species." (1833; translation in Daniel 1950:34). It is rather a cruel trick of fate that J.Boucher de Perthes should have received the

acclaim for the 'discovery' of the antiquity of man with artefactual evidence alone from the open sites of the Somme Valley. However, the support given by British quaternarists, such as H.J.Falconer and J.Prestwich, themselves primed by their own work at Brixham (Windmill Hill) Cave (Devon), may have finally tipped the balance of academic opinion in Boucher de Perthes' favour, thus salvaging a modicum of glory for the cave researchers. Once the biblical framework had been successfully challenged, there was no longer any available theory which might set a lower limit on the antiquity of man. Claims for Tertiary Man were made periodically right up until comparatively recent times when man's true tertiary ancestors and cousins were recognised. One of the most spectacular open-air finds was a complete skull, of remarkably modern type, recovered in 1866 from pliocene gold-bearing gravels at Calaveras (California). It was eventually realised that the find was a hoax, since there was unaltered stalagmite adhering to the skull and the Indian burials in nearby caves were well known and easily accessible (cf. Vayson de Pradenne 1932). In this case, it is difficult to judge whether the cave should be seen as the villain or the hero of the piece.

There has always been a great temptation in the study of palaeo-anthropology to give each new, and rare, hominid find distinct taxonomic status. Thus, some of the cave specimens were given unjustifiable names: Homo primigenius var. krapinensis Gorjanović-Kramberger (the sandstone rockshelter of Krapina, Yugoslavia); Homo calpicus Keith (Forbes' Quarry Cave, Gibraltar); Homo chapellensis Buttel-Reepen (La Chapelle-aux-Saints, Corrèze); Plesianthropus transvaalensis Broom (Sterkfontein, Transvaal). However, many taxa still retain at least an oblique reference to their original findspots: Homo sapiens rhodesiensis

Campbell (Broken Hill Cave, Zambia); Homo erectus pekinensis Weidenreich (Choukoutien Lower Cave, near Peking); Homo (sapiens) neanderthalensis King (Grotte Feldhofer, in the Neandertal near Düsseldorf); Australopithecus africanus Dart (Taung, Cape Province). It has even been suggested that certain cave specimens represent different racial stock: Cro-Magnon race (Quatrefages & Hamy 1882), Chancelade race (Testut 1889) and the 'negroid' Grimaldi race (Verneau 1906). The Catalogue of Fossil Hominids (Oakley et al. 1971, 1977) gives a more objective idea of the importance of caves to pleistocene palaeo-anthropology. 301 hominid findspots are listed for Europe, of which 231 (over 75%) are caves. Out of a total of 87 sites in Africa, 40 are caves, the prolific open sites such as those of the Hadar (Ethiopia), Koobi Fora (Kenya) and Olduvai (Tanzania), being reasonably balanced by rich cave sites such as Makapansgat Limeworks, Sterkfontein and Swartkrans (all in the Transvaal).

The discipline of prehistory, especially the study of the Palaeolithic, has probably received the greatest impetus from cave exploration. A recent bibliography (Schmider 1979) cites 582 major upper palaeolithic sites in France, of which 415 (over 70%) are caves or shelters. In Britain, over 65% of upper palaeolithic findspots are caves or shelters (Campbell 1977). Of the 454 lower, middle and upper palaeolithic sites from around the world singled out for special reference by Müller-Karpe (1966), 274 (over 60%) are caves or shelters. Among the many known upper palaeolithic occurrences of Cantabria, there is not a single true open-air site.

One cannot escape caves in palaeolithic and mesolithic research. The following caves and cave regions have all given their names, at one time or another, to archaeological cultures:

Abejas (Mexico), Amud (Syria), Antelias (Lebanon), Arene Candide (Italy), Asturias (Spain), Athlit (Israel), Aurignac (France), Bac-Son (Vietnam), Badegoule (France), Bayac (France), Capertee (Australia), Châteauneuf-les-Martigues (France), Châtelperron (France), Choukoutien (China), Creswell (Britain), Elmenteita (Kenya), Les Eyzies-de-Tayac (France), La Ferrassie (France), Goyet (Belgium), La Gravette (France), Hagfed ed-Dabba (Libya), Hagfet et Tera (Libya), Jerzmanovice (Poland), Kebara (Israel), Laugerie Haute (France), La Madeleine (France), Mas d'Azil (France), La Micoque (France), Montade (France), Montaigle (Belgium), Monte Circeo (Italy), Le Moustier (France), Mugharet el Emireh (Israel), Murzak-Koba (U.S.S.R.), Nachikufu (Zambia), Noailles (France), Oban (Britain), Périgord (France), Romanelli (Italy), La Salpêtrière (France), San Bernardino (Italy), Sauveterre-la-Lémance (France), Shan-Koba (U.S.S.R.), Szeleta (Hungary), Tabun (Israel), Trou-Magrite (Belgium), Wildkirchli (Switzerland), Wilton (South Africa), Yabroud (Syria), Zarzi (Iraq). When the name of a cave is too awkward, one can always resort to latin; thus Pech de l'Azé (France) gives Asinipodian and Gorge d'Enfer (France) gives Vallinfernalian! Even later periods of prehistory may follow suit: Covesea (Sculptor's Cave), in Morayshire, gives its name to a bronzeworking phase.

It is amusing to study the taxonomic rivalry in the early days of palaeolithic research, when so many new caves were being opened up. For instance, G. de Mortillet's scheme for the French Upper Palaeolithic originally contained Aurignacian, Laugerian and Magdalenian, all with eponymous cave/shelter sites. E.Piette, working in the Pyrenees, set up a series of cultures named after his own cave sites, such as Gourdan and Lorthet. However, the youngest culture in his sequence was called the

Arisian, after the river Arise which runs through the cave of Mas d'Azil. Of course, it did not take him long to see the error of his ways and change the name to Azilian. De Mortillet, meanwhile, let the side down by changing his Laugerian to Solutrean, even granted that Solutré is at the base of a limestone cliff. Piette was eventually rewarded for his loyalty as his Azilian replaced de Mortillet's Tourassian (itself named after a rockshelter).

Many artefact types are also named after cave sites. Some of the names quoted above occur again: Creswell point, Noailles burin, Gravette point, etc. Other sites boast only artefact types: Abri Audi knife (France), Cheddar point (Britain), Les Cottés point (France), Fell's Cave point (Chile), Font-Yves point (France), Gypsum Cave point (U.S.A.), El Khiam point (Jordan), Menton blade (Monaco), Olha flake (France), Peterfels point (Germany), Rgani knife (Poland), Sandia point (U.S.A.), Spy point (Belgium), Teyjat point (France), Les Vachons point (France). A problem arose in one Czech cave but luckily two names for the same site already existed; therefore, a Mladec point is a stone tool, whilst a Lautsch point is made of bone.

In the above review we have seen how important cave sites have been in the development of mammalian palaeontology, palaeo-anthropology and prehistory. In the early days, the whole community of quaternary specialists were enthusiastic about caves. Boyd Dawkins, archaeologist and palaeontologist, stated:

> The exploration of caves is rapidly becoming an important field of enquiry, and their contributions to our knowledge of the early history of the sojourn of men in Europe are daily increasing in value and in number. (1874:vii)

Geikie, a geologist with a general interest in the archaeological field, pushed the idea to the (familiar) extreme:

> Palaeolithic man was unquestionably a true troglodyte, the caves which he is known to have inhabited being very numerous. (1881:19)

The anthropologist Sir Arthur Keith put his proposition very clearly:

> In tracing the various kinds of men who lived in the Neolithic period, the open country, the river valleys, and the submerged land surfaces served us very well. When, however, we try to follow man beyond the bounds of the Neolithic period . . . , we must seek sequestered nooks where the earth keeps a more orderly register of events than in the turmoil of the flooded valleys. The ideal place we seek is a cave, particularly a limestone cave . . . . (1925: vol.1, p.66)

But this general enthusiasm was already on the wane when Keith wrote these words. Zeuner's textbook on the Pleistocene, written in 1959, contains only minor references to caves, despite the fact that Zeuner himself had excavated in several cave sites. West's Pleistocene Geology and Biology (1968) contains two short references to caves. The word 'cave' does not appear in Bowen's Quaternary Geology (1978). Except in France, where a strong tradition of cave research has so far survived, quaternarists in general, and quaternary geologists and micropalaeontologists in particular, are no longer greatly interested in caves. Even archaeologists, anthropologists and mammalian palaeontologists are seeking open sites more and more frequently. The phrases 'only of local importance', 'unclear stratigraphy', 'subject to mixing and pollution' or 'badly understood sedimentary processes' are being used all too often in connection with caves. Sparks and West have quite recently stated:

> Open habitation sites such as those considered above contrast with cave sites, where the stratigraphy is often far more complex and difficult, for the reason that the environment of deposition varies very much locally and so does the source of sediment, in the form of stalagmite, sediment from outside the cave and from the cave rocks. The regional environment is more difficult to study under these conditions, especially in the absence of well-preserved fossils of the kind we find associated with open sites and organic deposition. (1972:227)

Have quaternary studies outgrown cave sites?

1.2. The Development of Cave Sedimentology

In the above section, it was noted how rich caves may be in animal and human bones and in ancient artefacts. This tendency accounts for the success of cave sites in the early, object oriented phase of quaternary research. The recent loss of interest in caves is similarly accounted for by a desirable shift in emphasis, away from objects towards the concepts of context and association. It is clear that modern quaternarists have serious doubts concerning the quality and reliability of information from caves. The discipline most directly responsible for supplying stratigraphic and contextual data is that of cave sedimentology. This discipline has only lately achieved a measure of self-awareness; to trace its history, we must search the literature for often isolated cases of insight, involving excavators and theoreticians from allied disciplines.

The most striking and unusual of cave sediments is stalagmite. Not surprisingly, stalagmite formations occasioned the earliest enquiry concerning the nature of sedimentation in cave sites. The fact that many sites, such as Kent's Cavern (Devon) and Gailenreuth (Oberfranken, West Germany), were totally sealed by stalagmite before excavation was probably the best available evidence for the antiquity of the underlying deposits. Attempts were even made to measure the rate of stalagmite formation over surprisingly long study periods (cf. Boyd Dawkins 1874).

However, before the biblical model of geology was finally abandoned, the first coherent theory of cave sediments was proposed. This was an extension of the general Diluvial Theory, best expressed by Buckland (1823). Buckland observed a common

pattern in caves, with scree and other local sediments containing extant animal species often overlying exotic gravels containing extinct species. Naturally, the gravels were thought to be the result of the Flood.

As the more sophisticated Glacial Theory replaced the Diluvial Theory as the general geological model for the Quaternary, so cave sediments showing signs of exotic origin were referred to as 'drift' rather than 'diluvium'. Boyd Dawkins (1874) was the main proponent of the Glacial Theory in the cave situation in Britain. Of course, very little increase in understanding of actual cave sediments was achieved by this replacement of one a priori theory by another.

The next stage in the history of cave sedimentology, lasting from about 1880 to 1930, might be called the 'association phase'. This was the great period of cave excavation in Europe; few caves that were obvious at the surface escaped the picks and shovels, not to mention gunpowder, of the swarms of amateurs and not so professional quaternarists. Although much progress was made in the object oriented disciplines, such as prehistory, there was no coherent approach to the study of the sediments. Nevertheless, dotted all over Europe and often working in complete isolation, were talented amateurs whose enthusiasm led them to examine the sediments as well as the contained fossils. These were often comparatively uneducated men or local priests with little knowledge of international learning. The dedication that led them to spend most of their spare time down caves resulted in a quality of observation that has never again been equalled. Some of these 'little people', such as Peyrony and Breuil, eventually received the recognition they deserved, but many others left only their careful manuscripts (cf. reference to F.Moulin,

in Miskovsky 1974:25). The innovation they achieved lay in the recognition of association between different sediment parameters and between these parameters and contemporary faunas. It was found that climatic information could be retrieved in this way. By the end of this period, many researchers were beginning to see cave sediments as the result of complex climatic variation. Note that, although superficially similar to the earlier Glacial Theory as applied to caves, this new approach laid much greater stress upon data extracted from the cave sediments themselves. They did not totally escape a priori reasoning, however. Climatic determinism was slowly becoming the be-all and end-all of cave sedimentology.

To many observers, cave excavation during the 'association phase' appeared as total anarchy, just the activities of hordes of rather crass 'pot-hunters'. Many professional quaternarists, especially geologists, ceased to take an interest in the sort of deposits which attracted the object oriented disciplines. A communication gap began to form. For instance, no-one ever seemed to notice the possible implications that Clement Reid's theory of the formation of dry valleys on limestone (see p.29) might have for archaeological or palaeontological cave sites. By the turn of the century, a deep schism had developed between, on the one hand, prehistorians, palaeontologists and palaeo-anthropologists, interested in entrance facies deposits, and, on the other, professional spelaeologists and karst geomorphologists, interested in deep cave systems and limestone relief. Sadly, most of the geological expertise could be found, and can still be found, in the latter group.

In the 1920s and 30s, some researchers began to take the observation of cave sediment parameters a step further by

the introduction of simple quantitative methods. Amongst the leaders in this field were G.Riek (1934) and K.Utescher (1939) in Germany and G.A.Blanc (1921) in Italy. Many of the techniques were borrowed from other emerging disciplines, such as pedology, but some new methods, specifically designed to suit the cave situation, were also developed. The movement culminated in the publication of the article, "Uber Höhlensedimente", by R.Lais (1941), the first really thorough discussion of cave sediments. But the appearance of this article in Nazi Germany doomed the discipline to yet another schism. After the war, two separate schools of cave sedimentology emerged, one in Germany, based upon the work of Lais, and the other in France, drawing upon a long tradition of observation. Lais' article was eventually recognised in France by Vaufrey (1949) but, in seeking to heal the scars by publishing a French summary, Vaufrey unwittingly did Lais, and cave sedimentology, a great disservice. It is painfully obvious that none of the main French sedimentologists have ever read the original; both Miskovsky (1974) and Laville (1975) give incorrect references to Lais and, even in their bibliographies, immediately refer the reader to Vaufrey's summary. Much of the subtle detail in Lais' article was either not understood by Vaufrey or, more probably, was left out simply for reasons of space. When further opportunities for reconciliation presented themselves, the two schools still refused to communicate. E.Schmid, having been invited to work on the sediments of Lascaux, presented her findings for publication in France but the text appeared in German (Schmid 1968). As if in retaliation, the index to this volume has "E.SCHMID (Allemagne)", although Schmid was based at Bâle in Switzerland, and the typesetters seem to have done their very best to render the German even more obscure. The French archaeologist

and cave sedimentologist, J.-M. Le Tensorer, who was until recently based at Bordeaux, has now taken up a post at Bâle. It is to be hoped that he will be able to do something towards bringing the two schools together.

The work of many individual researchers since the war will be discussed in detail in later chapters.

The German School has found its main proponents in Lais (1941), Freund (1955) and Schmid (1958); its influence has spread eastwards (cf. Kukla & Ložek 1958; Vértes 1959; Brodar 1960; Brunnacker 1966; Osole 1968; Pelišek 1969; Lubin 1971), even reaching Israel (Schmid 1967). The German School is generally more aware of the importance of observational data and, although it has been guided by climatic determinism, there is some interest in the links between process and topography (cf. especially the work of Schmid). Quantification methods are usually quite simple. Brunnacker and Streit (1965) and Schmid (1969a) give useful overviews of this approach.

The younger French School owes much to Bonifay (1955), J.J.Blanc (1955), Escalon de Fonton (1969), Miskovsky (1970) and Laville (1973); their techniques have been taken up by Ek (1974) in Belgium and by Palmieri (1977) in Italy. Quantification is generally more complex but the approach is often rather rigid, with little consideration of the peculiarities of individual sites. Climatic determinism is usually the guiding principle. Laville (1976a) gives a summary of this approach.

Individuals working alongside but not totally integrated into either of the two main schools are often anglophones (eg. Shackley 1972; Farrand 1975; Goldberg 1979a; Butzer 1981). These researchers sometimes operate outside temperate Europe (cf. Shackley 1978; Farrand 1969; Goldberg 1979b; Butzer 1973).

They, and others like them, have often added new methods, developed within quaternary geology as a whole or in allied disciplines, to the classic battery of cave sedimentological techniques.

It is ironical that researchers working in Australia (cf. Frank 1971) or South Africa (cf. Brain 1967; Tankard & Schweitzer 1976) seem to be well acquainted with the major texts from both the German and French Schools, even though many of the techniques are not applicable due to the great differences in environment. Schools of cave sedimentology are developing in both these countries without compromising the spirit of international co-operation which happily prevails.

Owing to the writer's personal research experience, the present thesis is mainly concerned with European research in limestone caves and rockshelters. The reader may, however, be interested in comparing the methods developed to interpret sediments from caves in basalt (cf. Moser 1973) and sandstone (cf. Raynal 1975).

1.3. The Present Thesis

It is necessary to introduce an autobiographical note at this point in order that the reader may be fully aware of the aims of this thesis and the reasons which led me to formulate them. When, by pure coincidence, I took part in my first palaeolithic cave excavation in 1969, with J.Tixier at La Faurélie (Dordogne), I had not even started at university. I was, of course, totally unaware of the scope of cave research which I have tried to outline in the preceding sections. What is more important is that, when I began the research for this thesis in 1976, many things were still unclear. As it turned out, there were enormous gaps

in my basic knowledge of the subject, some of which I have not yet been able to fill to my own satisfaction.

I am by training, and strong inclination, an archaeologist, with special interest in the Upper Palaeolithic. However, following on from the study of geography at school, I took subsidiary courses in geology and physical geography whilst reading prehistoric archaeology at Edinburgh University. I became interested in the application of these subjects to archaeology and I further expanded my knowledge of the natural sciences with a diploma course in biology. Every summer, I returned to France to dig in cave and shelter sites.

When it came to choosing a subject for my M.A. dissertation I was guided by J.B.Campbell who was then lecturing at Edinburgh on, among other things, the Palaeolithic. He suggested that there was still work to be done in the cave sites of Creswell Crags (Derbyshire/Nottinghamshire). Although Dr. Campbell left Edinburgh before my final year, I undertook a project of section cleaning and the sinking of small exploratory trenches in several of the Creswell caves; it soon became clear that my dissertation would be concerned with the stratigraphic problems of these sites (Collcutt 1975). As part of the project I collected sediment samples, since I believed I could improve upon the simple granulometry already carried out by Campbell (1977). When it came to analysing these sediments, only two texts were available to me that dealt directly with archaeological deposits: Shackley (1972) and Cornwall's Soils for the Archaeologist (1958), still the best text of its kind in English, in my opinion. Other laboratory techniques I found in standard books on sedimentary petrology.

I learnt two important lessons during this work.

First, the amount of time needed for even the most basic laboratory analyses is staggering. Second, when you have followed all the instructions in the textbooks and come up with neat columns of figures and pleasing graphs, it is far from clear what you should do with them. Much too late to remedy the situation, I came to a vague realisation that I had not collected enough observational data on site and that I had not set myself specific problems to be solved by suitable laboratory analysis. Since that time, several students, even up to doctoral level, have come to me in exactly the same predicament. If this cookbook approach has been adopted, there is often no useful advice to be given and much painstaking work will have been wasted. For my part, only a few relevant facts could be salvaged from my practical work at Creswell.

From Edinburgh, I went to Bordeaux University, attracted by the wide range of environmental courses that were offered along with the main courses on the Palaeolithic. I read H.Laville's thesis, Climatologie et chronologie du Paléolithique en Périgord: étude sédimentologique de dépôts en grottes et sous abris (1973). Here was a methodology, simple yet inclusive, that would solve all the problems I had encountered at Creswell. I attended Laville's lectures and worked in the sediment laboratory. Although I came to have some reservations concerning certain techniques, I believed that Laville's approach was generally valid and that it should be tried in British sites.

I applied to do doctoral research at Oxford. My plans were simple: I would select a region, probably South-West England; I would examine known cave sites and perhaps discover one or two new ones; and then I would set up a regional chronostratigraphy, based on climatic inference , that would put

palaeolithic research in the area on a much firmer footing.

During the summer of 1976, my wife and I visited the great majority of known palaeolithic cave sites in my chosen area, together with many purely palaeontological sites. We were fortunate to meet many local researchers, perhaps the most encouraging being the late Professor E.K.Tratman. We discovered one new upper palaeolithic site (Pixie's Hole, Devon), which we have excavated over the years. However, it became clear that not many sites were available for the sort of research I had in mind. Many of the caves referenced in the literature had been destroyed by quarrying or were totally empty, ransacked by previous excavators not always, it is sad to say, in the very distant past. My plans for a regional chronostratigraphy now seemed threatened.

Further problems awaited me at Oxford. Despite the persistent, and much appreciated, efforts of my supervisor, Dr.D.A.Roe, the laboratory equipment I needed was not immediately available. As it happens, I now consider this to be the greatest stroke of luck, as it prevented me from trying to muddle through with my original project. I spent a full year following up all the literature on cave sedimentology that I had only recently discovered, starting with the French School but soon expanding to the German School and further afield. My work left me with a sense of gloom. It was obvious that not enough good sites existed for a regional chronostratigraphy. It was also clear that Laville's techniques were far from satisfactory on theoretical grounds. Indeed, I became hypercritical to the point where I began to wonder if anybody knew anything about cave sediments at all, so contradictory were the reports I had been reading. I published my pessimistic views, first as a short note (Collcutt

1977) and then in greater detail (Collcutt 1979a).

The reaction to my publication surprised me. Few cave sedimentologists reacted at all, but those few did so in no uncertain terms. For instance, I learnt that G.J.Kukla had taken exception, not unreasonably, to my rather unkind treatment of the work of L.Vértes in my 1977 paper. However, many quaternarists not directly involved with cave sediments indicated to me that my publications confirmed their own views and that they would continue to be suspicious of any sort of information, not just geological, derived from cave sites. I was, to put it mildly, chagrined. My efforts seemed to have had exactly the opposite effect to the one intended. Rather than stimulating interest in the study of caves, my papers were being used, at least in some quarters, to help justify the dismissal of such sites. I realised that my statement had been too negative. I was, however, at a loss as to what to do about this, since I was still unable to propose a coherent alternative to the established methodologies that, at least to my mind, had proved ineffective.

I have always spent as much time as possible in the field. To date, I have been involved, in one capacity or another and for varying lengths of time, in the examination of over sixty cave and shelter sites. I have come to believe very strongly in observation and the recognition of exact processes. Laboratory analysis is merely a tool, and an inappropriate one at that in the hands of someone who lacks the experience to ask the 'right questions'. The 'right questions' are not always obvious but they are often rather simple. Traditionally, the archaeologist's first question to the sedimentologist will be: "Under what climatic conditions did this deposit form?". This is a valid question but it will very rarely be a 'right question' from the

sedimentologist's point of view because it is much too complex. It would be like the sedimentologist handing the archaeologist a flint artefact and asking: "under what socio-economic conditions did this form?". There are much more basic questions to be answered first and, in many cases, the more complex problems will remain unsolved through lack of relevant data. At least for the foreseeable future, climatic determinism will continue to be a dangerous illusion.

This thesis is therefore about the 'right questions'. It should not be assumed that the 'right answers' will necessarily be forthcoming; I am by no means convinced that I have accounted for a majority of the questions. I have gone right back to fundamental principles and suggested a methodology that has an almost philosophical basis, as independent of specific analytical techniques as possible.

It is a moot point just how far this methodology may be considered original. Some aspects have been taken, en bloc, from the published work of others. Some aspects result from the combination of various ideas expressed by others. When information on a given topic was not available from cave sites - an extremely frequent occurrence - I have gone to the general sedimentological literature. Since I have erred on the side of the meticulous with respect to sources, often using direct quotes rather than simple references, the reader may judge the question of originality for himself. Quotations are normally given in the original language. Where necessary, translations have been added. The guiding principle in producing these translations has been exact communication of content rather than literary style. If I feel that I do not fully understand a passage, or that a passage contains a significant level of ambiguity, I have tried to indicate

the uncertainty as clearly as possible. To my knowledge, the approach advocated here has not before been used in archaeological cave sites. However, I personally find it reassuring that, in the last decade, there has been an increasingly confident move in sedimentology as a whole, away from laboratory analysis towards direct observation of depositional processes and the resulting sedimentary structures.

In that I believe that I can now propose some positive aspects of cave sedimentology, I feel justified in returning to the criticism of earlier work by others. This is a destructive task, in which I take no pleasure, but it is strictly necessary if I am to make my point. Naturally, I tend to criticise those researchers whose work I know best. In so doing, I perhaps create the impression that I am singling them out for special attention. This is not the case. Also, I concentrate upon what I see as shortcomings and often leave what is obviously valuable research to speak for itself. I am interested in theory; it is not always possible to generalise from an individual statement which one holds to be true, but it is always possible to isolate a statement which is contrary to a particular theory. This results here in a rather jaundiced text and I apologise to the researchers concerned, especially to MM. Laville and Miskovsky. With one notable exception (Lwoff 1974), I have never read a paper on cave sediments which did not leave me with the impression that my understanding of the subject had been improved.

Although incomplete in many respects, this thesis is essentially a 'textbook'. There are masses of synthetic material. My own practical work is presented by way of illustration of some of the theoretical considerations that have arisen; the reader would be greatly mistaken if he were to evaluate this material

as if it were meant to represent discrete site reports. One major deficiency in the thesis as a whole is the poor photographic coverage of many of the sedimentary structures and geometries described in the text. Apart from factors such as cost and the extreme technical difficulties involved in taking close-up photographs in caves, the reason for this is that, until recently, I had not formulated the conscious aims which now require such documentation (see p.1229); there is no precedent for detailed photographic coverage in cave sedimentology.

I have said that I consider myself to be an archaeologist. I hope that this thesis will be of interest and use to archaeologists. However, the text is overwhelmingly geological in content, with only brief reference to mainstream archaeological matters. My justification for this is that if we cannot re-establish confidence in cave sites at the stratigraphic and contextual level, the support of more and more quaternarists will be withdrawn. It will be palaeolithic archaeology that will suffer the most immediate loss if this should happen.

My overall thesis is therefore that, at present, much is wrong with cave sedimentology; that at least partial remedies exist for these ailments; and that quaternarists, and especially archaeologists, would be greatly mistaken if they were to ignore the contribution, actual or potential, of cave sites.

1.4. A Basic Framework

If caves appear to be unsatisfactory for innovative quaternary research, it is perhaps at least partly because we are asking too much of them in some areas and not making full use of their potential in others. It will therefore be necessary to

examine the general nature of caves and cave deposits. This section is deliberately included here, in the 'Introduction', since it complements what has already been said above. The four sections of this chapter together provide a perspective which will guide the more detailed discussion presented in the body of the text.

There is often, if not always, a massive time gap between the formation of a cave and the deposition of the oldest sediments available for study in the present. The earlier stages of spelaeogenesis are therefore only of interest to us in as much as the past history of a cavity will have an effect upon patterns of sedimentation. Schmid (1969a) has suggested an empirical division of caves into two classes: endogene caves, that penetrate far into the limestone massif, and exogene caves, that are shallow features at the surface of the massif. Furthermore, as the names suggest, endogene caves will owe much to processes acting within the massif, whilst exogene caves will result primarily from processes which encroach marginally upon the massif from the outside. This is a useful division for the purposes of sedimentology but it is somewhat arbitrary and has no time dimension. A more complex model, yet still very simplified from the point of view of karst geomorphology, will be helpful.

Most larger caves are the direct result of the need to drain meteorological water from and through limestone, a soluble but generally impervious rock type. Consider a region in which a large river drains a rock succession, with limestone overlying an insoluble rock, such as sandstone, on a sloping contact; moving across the landscape, we start on the sandstone, come to the limestone and then descend into the main valley. Tributary streams run overland on the sandstone but pass underground upon reaching the limestone; they eventually emerge near the bottom of the main

valley just before joining the river. The limestone is drained through a series of fissures and cave passages. Some deep parts of the system will be permanently and totally filled with water; this is said to be the phreatic zone. Higher parts of the system may have streams, with free surfaces in contact with air, which may be more or less intermittent, or just generally air filled cavities with water dripping and trickling through them; this is said to be the vadose zone. Within the vadose zone, where water is flowing quite regularly in streams, the cave may be said to be active, since these passages play a major role in the drainage of the limestone. Where there are no streams and only comparatively minor transfer of water, the cave may be said to be inactive. Note that movement of water also occurs even deep within the phreatic zone. Because water is so concentrated, within solutional cavities and structural discontinuities (fissures, faults, joints, etc.), rather than dispersed throughout pores as in many rocks and sediments, the concept of a water table is difficult to apply. The top of the phreatic zone may be considered to be the water table (the word 'phreatic' having its root in the Greek for a 'well'), but this surface will be much more irregular and stepped than the water table in a porous medium. Nevertheless, the drainage within limestone is still part of the regional pattern and the cave waters may be said to be graded to the base level of the river in the main valley.

If a time dimension is added to this model, erosion of the landscape will come into play. One aspect of erosion will be the deepening of the main river valley. As the water table drops, some originally phreatic passages will become integrated into the vadose zone, some areas within the vadose zone will change from active to inactive conditions, and some parts of the inactive

system will become so isolated from the main drainage pattern that they may be said to be 'fossil'. New passages may be added at the base of the phreatic zone but the mechanisms involved need not concern us (cf. Renault 1967, 1968, 1969; Bögli 1980). Retreat of the valley sides will slowly destroy sections of cave passage, as will a general surface lowering of the limestone 'plateau' above the valley.

Sediments laid down in the active vadose zone, and even sometimes in the phreatic zone, will be of interest to the object-oriented disciplines of quaternary studies. However, bones or artefacts contained in these sediments will be in derived, secondary contexts. It is the inactive or very intermittently active sections that may contain primary context assemblages. We will return to this point below, but it is important to envisage the span of time between abandonment of a section by active cave streams and destruction of that section by erosion from the surface. Both ancient events, such as the habitation of a cave by man, and investigation of the site by a quaternarist must take place within this time span.

We may now move on to a more detailed model. Fig.1 represents a theoretical landscape that has been evolving over a long period of time. Some of the stages in this evolution are clearly discernible. The oldest extant <u>influent</u> cave is at G*; the tributary stream ran over the cap rock C until it was able to pierce this impermeable layer. Further erosion caused the point of entry to regress upstream, first to G** and then to D, the currently active influent. Purely local catchment is represented at I; a <u>doline</u> is a closed depression, usually of quite small dimensions (at least in Britain), with water percolating through fissures, rubble and sediments rather than running freely into a

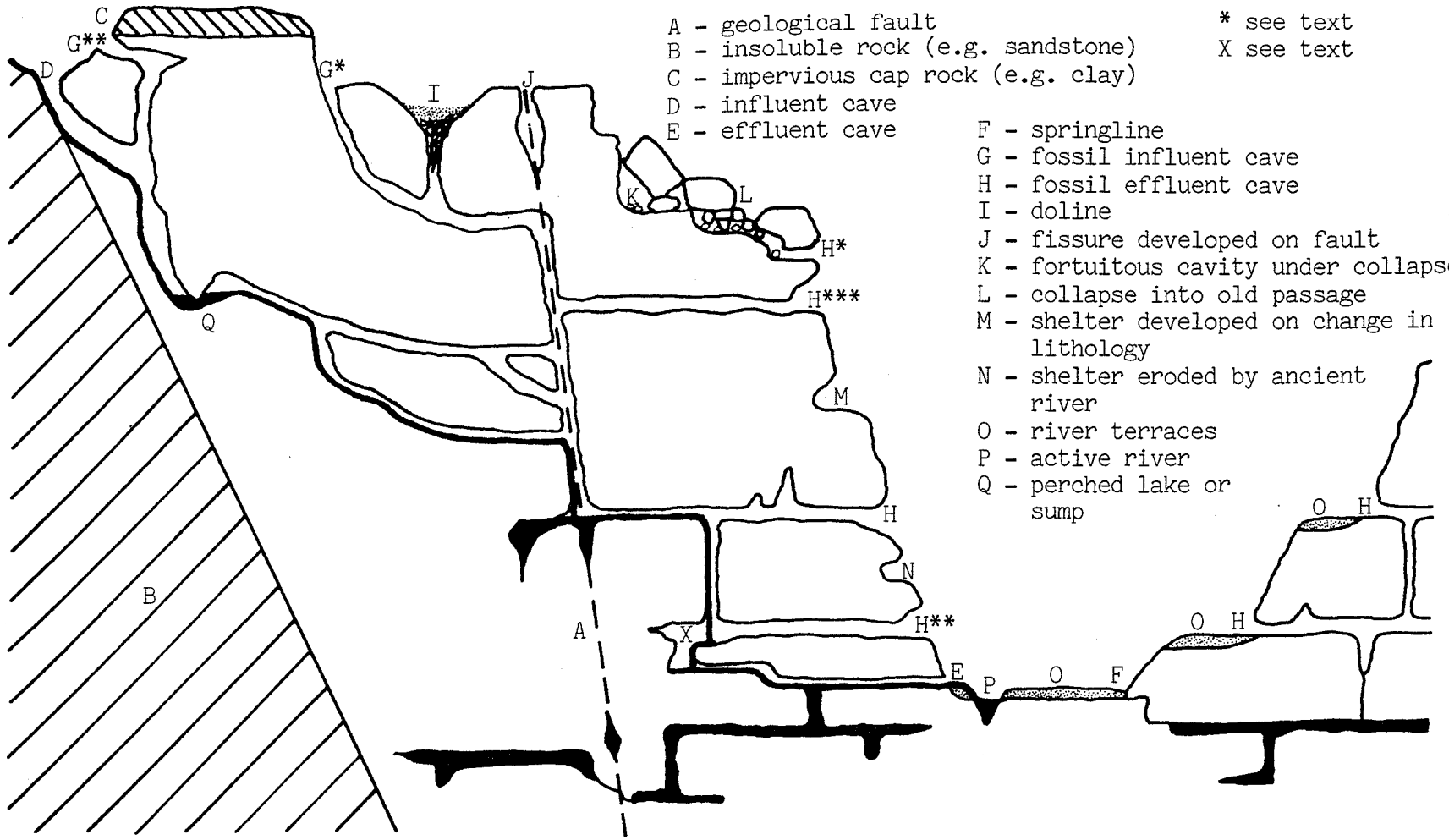


Fig.1 Theoretical Landscape With Caves

well defined cave mouth. The effluent caves, the abandoned ones marked H and the active one E, represent progressive readjustments of the cave system to valley incision and lowering of the water table. The link between cave systems and regional hydrology is further stressed by the coupling of the river terraces, marked O, and the inactive effluent caves on the far side of the valley. This pattern of drainage within the limestone produces caves of Schmid's endogene type.

Other types of shallow cave and shelter are also shown in fig.1. J is a fissure, developed on the major fault A; from that part of it which remains, it would appear that it has never been integrated into the main drainage system. K is a fortuitous cavity, roofed over by fallen blocks; such cavities are normally quite small, although there are some notable exceptions. M is a shelter developed upon a change in lithology in the limestone which is causing preferential erosion (eg. by frost fracturing) at this point; if the limestone is more or less horizontally bedded, such shelters may run continuously for a considerable distance along a valley side. N is a shelter actually eroded by a meander of the main river as it bit into the rockface at an earlier stage in its evolution. Laville (1975) illustrates and discusses shelters of types M and N, which are extremely important archaeologically in the Périgord region of France. These shallow features are examples of Schmid's exogene type.

Apart from showing the general evolution of cave systems in response to falling base level, fig.1 contains much more information of great interest to the cave sedimentologist. Note how the inactive influent cave G** has survived rather longer than usual, due to the protection of the impervious cap rock C; it may contain undisturbed material of considerable antiquity.

The section L to H* is probably even older but the 'entrances' have been gradually approaching one another through time; in fact, this section is very close to total obliteration. The shelter N will soon be destroyed by cliff recession, since the process by which it was formed has ceased to operate. On the other hand, the shelter M, although it too is being constantly attacked by cliff recession, will survive as long as the lithological change activates the erosion process. The river terraces O provide a means by which the local cave sequences can be correlated with other sites in the region.

So far, discussion has centred upon a system graded to a more or less constantly dropping base level. Rather than a river, a lake or the sea may provide a base level. It is a familiar image that, during past interglacials, sea level was often higher than it is today. Less familiar is the concept that lake and river levels may also have been subjected to a temporary reversal of the normal downcutting trend. In many parts of Europe during pleistocene cold periods, massive quantities of ice and moraine were available either to dam valleys, and thus raise lake and river levels, or to raise the effective level of the whole valley floor. Rivers would remain 'perched' until the ice melted or until fluvial erosion had cut down through the moraine. It is obvious that if we fill our valley in fig.1 with impervious moraine up to the level of H**, the local water table will probably rise, parts of the cave system that were formally inactive may be flooded and return to active or even phreatic conditions, and the once inactive cave H** may be <u>reactivated</u> as the current effluent. Old sediments, together with any contained bones or artefacts, that might have been in the H** passage could be disturbed or totally removed during this reactivation.

There are some limits to the amount of ice or moraine available to choke a whole valley. We might assume that 'fossil' sections high on the valley sides would be less vulnerable to reactivation in this way. This is probably true but there are other ways in which previously 'fossil' sections may be reactivated, with the same disruptive results.

The simplest way in which this may happen is by collapse deep within the system. Returning to fig.1, if the roof of the chamber at X collapsed, the stream, or at least part of the water, might be forced to flow out through the exit at H** until the blockage could be removed. Note how, although changes in base level (fluvial, lacustrine or marine) may be ultimately determined by climatic factors, the collapse of part of a cave system is the result of a complex combination of events, sometimes very far in the past (the 'history' of the cave). Although a climatic, or more probably a meteorological condition (eg. flooding) may trigger the collapse event, the event will not be predictable from climatic data, nor will the recognition of the event necessarily indicate particular climatic conditions.

Reactivation of another type, this time with strong but rather generalised climatic control, seems to be particularly common in British caves. However, it was probably very important in all areas which were once glaciated or which were within the permafrost zone at some time during the Pleistocene. Reference was made in section 1.2. to Clement Reid's theory concerning the origin of dry valleys in limestone, and specifically chalk, landscapes (Reid 1887). Chalk is so porous and permeable that water rarely runs at the surface. How then could valleys be formed that had all the signs of a subaerial erosive origin? Reid proposed that, during the Pleistocene, permafrost conditions would

render the chalk impervious, thus initiating surface drainage. Warwick, whilst discussing such dry valleys and Reid's theory, notes that dry valleys "contain no streams except occasionally after a severe frost followed by a quick thaw, as in 1947 [on the English chalklands]."(1976:69). Other mechanisms have also been proposed for the formation of dry valleys (cf. Sparks 1972; Warwick 1976) but this is not important for our purposes. What we need to know is what happens inside a cave system if input of water from the surface is reduced or almost completely stopped. Permafrost itself might cause this, by freezing of sediments covering the innumerable fissures on limestone plateaux or within dolines, but it is not the only mechanism. Mass movement, a very common set of processes in the periglacial environment and by no means unknown under more temporate climates, could block normally influent caves. Glacier ice and associated moraine could seal limestone outcrops as well as valley floors. If flow in a cave system is significantly reduced over long periods, the stream may not be able to keep its channel open. Masses of sediment may be introduced into the system during spring melts but, without a constant and efficient stream, much of this sediment will accumulate in the system. Slowly, the former streamways will become choked. When the climate eventually returns to warmer and effectively more humid conditions, streams may have to re-excavate their former channels. In the meantime, all or part of the input water may be forced to exit through previously 'fossil', high level passages. Even H*** in fig.1 could be within the reach of this process.

The importance of reactivation cannot be overestimated. A stream running in a formerly dry passage will often cause much disruption of older deposits. Much sediment will be in transit

at least seasonally at such times and water flow may give way to the flow of highly destructive water/sediment mixes. If a passage is totally blocked by sediments or stalagmite, a head of water may build up behind the dam. When the dam eventually ruptures, the passage may be totally scoured of all but the most indurated deposits, a process often referred to as flushing. Sudden floods due to excessive rainfall or melt-off may have a similar effect (cf. the analysis of response to a flood situation in Smith et al. 1976, especially Fig.6.14 and relevant text); as was noted above, sudden flooding is usually a meteorological rather than a climatic phenomenon. Although it may take tens, or even hundreds, of thousands of years for sediments to build up, it may take only hours to remove or disrupt them. Many cave systems are known to have formed several million years ago; reactivation is the reason why caves almost never contain sedimentary sequences starting from the time stream flow first waned. Traces of very old deposits may often be found at the bottom of a cave sequence. That they are usually unstratified and hopelessly mixed, or at the very least deeply dissected, is a testimony to the near inevitability of reactivation.

The concept of reactivation can be integrated into a general model of energy levels in caves. We have seen how cavities may be formed by comparatively high energy processes, such as efficient drainage or 'cataclysmic' roof falls. When the erosive agent has migrated, or the 'cataclysm' has passed, the cavity lapses into a period of relative quiescence, it becomes inactive or even 'fossil'. At any time in the future, energetic conditions may be renewed (reactivation, tectonic disturbance, etc.), only to be followed by another period of quiescence. The history of a cave and its sediments can thus be seen to be broadly cyclic. Caves are rather

different from most geomorphological entities, owing to the fact that they are more or less surrounded by the rigid structure of the rock mass. During periods of quiescence, the external processes of weathering and erosion, those processes which constantly degrade the Earth's surface, may be greatly attenuated by the protecting rock. In deep fossil caves, energy levels may fall to values unknown other than in ocean basins and deep lakes, situations in which masses of water rather than rock are the buffer. But when a high energy process penetrates the cave, the rock prevents rapid dissipation of the energy and channels destructive power through the cavity. It can therefore be seen that caves are generally subject to energy cycles of extreme amplitude but relatively low frequency, the high peaks occupying only a very small percentage of the time involved. The nearer one approaches the surface, the more conditions approximate those of the surrounding geographical area. Nevertheless, even the smallest rockshelters tend to show this pattern on a greatly reduced scale.

Let us return to fig.1 and examine some of the caves with respect to their position within the energy network of the area. We have already seen that shelter N will certainly be destroyed by cliff recession. However, given its isolated position in the local topography, it is rather unlikely that it has suffered major disturbance since the river left it high and dry. The section L to H* has also been isolated from the main drainage system for a considerable time. This is not true of the 'fossil' effluent caves above and below shelter N; we might expect them to have suffered comparatively frequent disturbance in the not too distant past. The cavity K is well isolated from the drainage system but its topographic position, and the fact that its roof is not solid rock, suggest that it is only a relatively transient feature.

Shelter M is constantly migrating and we would not expect to find very old deposits in it. In fact, since its roof or walls are developed in a structurally weak zone of the limestone, periodic collapse may mean that it is only available to trap sediments of interest to us for short and irregularly spaced intervals. The lake or sump at Q may appear to be a perfect sediment trap. However, this feature has obviously been a part of the main drainage channel for a very long period; every time water flow fluctuates, the sediments in Q are at risk. Given a period of evolution a little beyond the stage shown in fig.1, the cave at G** may achieve long term stability. A cave, totally filled with sediments, within an isolated outcrop capped by insoluble and impermeable rock, lying well above the local water table and with its entrances sealed under talus incorporated in stable slope deposits, may well survive for several million years.

Let us now reconsider the widely held view that cave sites have major shortcomings when compared with open sites (see p.9). The criticisms fall under these main headings: complex stratigraphy, due to varying local topography and varyied sediment sources; poorly understood sedimentary processes; lack of organic deposition with well preserved fossils; isolation from regional environments.

Cave sediments are often divided into two main classes: autochthonous sediments, derived from the surrounding limestone, and allochthonous sediments, brought in from outside the cave. This is a useful division as long as the term 'sediment' is understood to mean 'component of a sedimentary unit'. It is obvious that most sedimentary units within a cave, or anywhere else for that matter, will contain mixtures of material from many sources. Only on the occasions when a high energy process penetrates the cave (eg. fluvial, aeolian or marine processes) will

a true allochthonous sediment body be created. Similarly, even stalagmite usually contains some detritus which may have been introduced from outside. The sedimentary processes operative within a cave are, of course, the same processes that operate everywhere; it is the expression of these processes, the 'style' of the sediments, that may be slightly different. The fact that we may not yet fully understand these expressions is not a shortcoming of the cave sites themselves. There is absolutely no evidence to suggest that any present difficulties are intrinsically insuperable. By lack of organic deposition, critics presumably mean lack of plant material. This is generally, but by no means always, true, because cave sediments are often rather well aerated. Note that it is oxidation, not alkaline conditions, which destroys organic tissue; a few cave deposits are sufficiently waterlogged or rich in colloids to provide a suitable environment for preservation. On the other hand, cave sites usually contain very much greater quantities of bone material than do open sites, this time because of their alkalinity and the availability of solutes which speed fossilisation (lithification). Cave sites, as a rule, are not isolated from the regional environment; this misconception is due merely to the way cave sites have been studied in the past. In as much as environment and climate are recognisable in any sedimentary body of a given scale, such information can be retrieved from caves. Even the spatial isolation of many cave sequences may be more apparent than real, resulting from the lack of excavations extended beyond the caves themselves. The survival of a very old deposit naturally depends upon its isolation from the effects of subsequent environments, whether in a cave or in an open site.

In order to gain a proper perspective in this discussion,

let us briefly consider types of open site which critics of cave sites often feel are so much more informative. When discussing biogenic sediments, West states:

> Several criteria are used in a genetic classification:
> 1. Position of the deposit in relation to the free water surface or groundwater. (. . .)
> 2. Manner of deposition. Biogenic deposits are autochthonous (or sedentary) where the plant detritus is deposited where the plant is growing. (. . .)
> When the plant detritus is transported and then deposited, . . . , an allochthonous (or sedimentary) deposit is formed. (. . .)
> 3. Supply of nutrients to the plant communities producing the deposits. (. . .)
> 4. Topographical and drainage conditions. (. . .)
> (1977:55-7)

If these are not the same factors which influence cave sediments, they are at least very close analogues.

Sparks and West state:

> Lake sediments receive macroscopic plant remains from water plants growing in the lake and from plants growing around the lake. They also receive pollen and spores from the atmosphere and from streams draining into the lake. Marine sediments may similarly contain macroscopic and microscopic plant fossils. Peat deposits contain the remains of plants growing at the surface of bog or fen, the pollen and spores of these plants as well as pollen and spores received from the atmosphere. (1972:165-6)

> Errors of interpretation may arise in several ways. Weathering of sediments causes differential destruction of pollen grains. (. . .)
> The presence of pollen derived by erosion of older sediments and re-introduced into a sediment may interfere with the interpretation. Here a consideration of the origin of pollen in a sediment will be of help. Fig.6.5 [a detailed and informative diagram] shows the various ways pollen may become incorporated into a sediment. Derived pollen is likely to be present in lacustrine and marine deposits receiving drainage water containing pollen grains removed from older sediments. Some estimate of what is derived may be obtained by pollen analysis of these older sediments through which the water is draining. On the other hand, in raised bogs . . . and blanket bogs . . . , where the growth of the peat surface is dependent on high rainfall and the surface of growth is above the regional water table, all pollen present comes from the local plants or by more distant air transport from regional vegetation so that derived pollen is hardly likely to be present.
> Each site of pollen analytical study has its own particular conditions of pollen receipt. The determination of local and regional components through knowledge of the more locally derived macroscopic plant remains will aid in the interpretation of vegetational history. (ibid., pp.172-4)

It is not difficult to see the precise parallels between this discussion and that concerning caves and cave sediments given above. Similar analogies could be drawn with the factors which

influence the interpretation of coleopteran or molluscan remains. Even if autochthonous peat is not acceptable as a true analogue of stalagmite, open sites near springs have tufas and lakes have calcareous marls.

It is clear to the present author that all quaternary specialists face very similar problems, whatever type of site they choose; the picture is never simple. However, this complexity can be made to work for us. As palaeobotanists have often demonstrated, the idiosyncrasies of individual sites, once understood through careful study, can turn what was initially bewildering complexity into a wealth of detailed information about the immediate environment of a site, information which will be invaluable to our appreciation of the relationships between objects (biological, geological or archaeological) found within the site. Without such basic information, we have little hope of understanding the synthetic topics (ecological, economic, cultural, etc.) of interest in archaeology.

We may now turn to the specific subject of archaeological and palaeontological cave sites. In section 1.1. it was noted how there has been a strong bias towards cave sites in the object--oriented disciplines of quaternary studies. It might be considered that this bias is merely a historical accident, a whim of the early excavators which has left the image of the 'cave man' or 'cave bear' firmly implanted in our modern folklore. On closer inspection, it can be seen that this bias (and bias it most certainly is) is quite logical and, with respect to the healthy development of the disciplines, most fortunate.

It would be rash to suggest that palaeolithic man preferred caves and shelters to open-air sites. However, between erosional events, cave sites do provide protection from the weather and

from predators, so that it would be most surprising if they had not been used by man at all. The same argument applies to den carnivores, bats, birds of prey, etc. Having chosen a cave, the occupant is forced by the morphology of the cavity to confine his activities to a relatively small area. If an unstable, transient cave or shelter has been chosen, it will shortly be destroyed; there will be little time for a long series of occupation layers to accumulate. If, however, a stable cavity is occupied once, there is no reason why it should not be occupied several times. Indeed, the probability is that such a stable feature in a changing landscape will almost certainly be visited many times during its history, each visit contributing material to be distributed over approximately the same small area.

It is this argument, rather than the small scale of occupation, which appears to the author to indicate low demographic levels during the British Upper Paleaolithic. Small scale could theoretically be due to a strong cultural bias towards open-air sites for some reason; it is entirely possible that many large open sites exist that have not yet been found. However, if the caves were available, yet even very small scale multi-level sites are rare and many caves totally unoccupied, the logical conclusion is that man was not there in large enough numbers to occupy them more regularly. Compare the situation in the Périgord, where practically every available site was occupied many times.

Thus, debris from animal and human occupation, together with material introduced by geological processes, will accumulate with time in the low energy trap of the cave. Eventually, this debris will be destroyed, mixed, removed or altered by a major erosional event. We have seen how the cave's position in the local geomorphological setting will affect the likelihood of an

erosional event.

The archaeologist or palaeontologist, once he has realised that caves and shelters often contain cultural or faunal debris, will consciously seek out such sites. He will have little difficulty in finding those sites which are visible on the surface. The fact that caves are comparatively stable landmarks in a changing landscape therefore accounts both for their high probability of selection as living sites in the past and for their selection as excavation sites in the present. However, the group of sites visible at the surface at some time in the past will not be the same as that visible in the present. Using the concepts set out above, the researcher should be able to judge the likelihood of a site being of interest to him. Very young sites (unstable, transient cavities and cavities only recently abandoned by main drainage) may be sterile or contain only very recent material. In older sites, still potentially within the main drainage system, the increasing probability of a major erosional event with increasing time since the last such event (cf. general catastrophe theory) will ensure that most archaeological or palaeontological finds will be comparatively recent. Very early sites will only survive under exceptional circumstances; they will nearly always be more or less invisible on the surface. Similarly, older transient caves, with short stratigraphies, will also be invisible. Such sites will only be found by chance or due to natural exhumation, events that are improbable at any one point in time, thus reinforcing the rarity of known early cave sites. Only conscious analysis of the landscape, coupled with remote sensing procedures (eg. resistivity survey), can better the chances of recovering old sites.

The observed bias towards palaeolithic, or even later palaeolithic, cave and shelter sites therefore appears quite logical.

We are aware of it, yet there is little we can do to change the situation, as all types of earlier sites and later open-air sites will nearly always be invisible on the surface. More continuous natural erosion, together with large scale commercial excavation, make it more probable that early sites will come to light in the open rather than in caves, although vastly accelerated limestone quarrying in Europe is now increasing the incidence of discovery (and rapid destruction) of early cave sites. That this bias towards caves has proved most fortunate for the development of at least upper palaeolithic archaeology is shown by the relative ease with which cultural sequences have been established in areas with abundant caves, compared with the uncertainty remaining in areas where caves are rare or still unstudied. Caves generally have rather slow sedimentation rates during periods of quiescence and the sediment is often comparatively coarse; this makes extremely fine stratigraphic control (i.e. on a scale of decades) very difficult. Couple this with the likelihood of frequent animal and human occupation and it can be seen that caves are usually best suited to a 'vertical' approach, with an accent upon change through time. Open-air sites, with higher sedimentation rates, finer sediments and usually lower potential for multiple occupation, are better suited to a 'horizontal' approach, with an accent upon detailed reconstruction of environment and individual short term events.

PART I

THEORY

2. THE CAVE ENVIRONMENT

2.1. General

In section 1.4. above, the subject of cave environment was introduced during the general discussion of energy levels in caves. It was seen how caves fit into the hydrological system of a region and how the surrounding limestone may buffer, to a greater or lesser extent, the direct effect of climate. As one moves from exterior to subterranean localities - from valley side, to rockface, to rockshelter, to cave entrance, to cave passage, to deep cave sections - the overall environment changes extremely rapidly in just about every parameter which would normally be used to express environment (temperature, humidity, air movement, atmospheric composition, light levels, biotic factors, etc.). There is no single cave environment; there is a rapid succession of micro-environments.

Before continuing, a clear distinction must be made between the concept of environment, as employed in this chapter, and the concept of 'environment of deposition'. Although the latter is part of the cave environment as a whole, it is a very much more restricted concept, applying only to such factors as the nature and speed of transporting media or the configuration of the area of deposition, factors which we are confident we can observe or deduce accurately from the sediments (cf. Chapter 3). In this chapter, environment is discussed in its wider sense,

involving aspects of climate, weather, ecology, physical setting, and the like.

The difficulty with cave micro-environments is that, although we can apply much general knowledge and discuss the subject on the theoretical level, there is no available synthesis which will serve as a tool that may be used to interpret ancient deposits. Enormous quantities of basic data have certainly been collected, especially in the biological sciences, but these data are scattered throughout the literature of many countries and it would take a concerted effort to produce an overview. The author has come across detailed information concerning how a particular species of bat or bird of prey might choose its roost within a specific site, apparently as a function of temperature, humidity, air movement or light level (cf. Vandel 1965; Daan & Wichers 1968). But how far can such individual observations be generalised in order to allow micro-environmental inference to be drawn from the occurrence of undisturbed bat skeletons or owl pellets? Similarly, we have a reasonable understanding of the chemistry of speleothems (carbonate precipitates including stalagmite) and we know which environmental parameters must be involved, but we cannot yet draw any detailed inference from the composition, structure or form of ancient examples (but see Chapter 13). Theoretical models can only go so far before the system becomes too complex to provide useful information. What is needed now is a body of empirical data to furnish the theoretical models with scalar estimates of how important given parameters may be under real conditions.

The following sections are composed almost entirely of theoretical discussion. Only a very few cases of direct observation are available and these are not linked with studies

of sedimentary processes.

Most texts on cave sedimentology have a warning section, like the one given by Butzer:

> Each cave must be assessed and interpreted individually, since cave ventilation, temperature amplitudes, and water regime vary considerably even with local areas. Even within a single cave there are significant differences of microenvironment at the entrance, in the interior, and at different elevations within the deep passages and fissures. (1971:213)

If anything, such a warning should be considered to be a considerable understatement.

2.2. Cave Meteorology

The subject of cave meteorology is best approached through theoretical models of the relationships between the three main atmospheric variables: air movement, temperature and humidity. The starting point for the following discussion is provided by the work of Wigley and Brown (1976); the detailed physics underlying these models are set out in Wigley and Brown (1971).

The deep cave environment is not usually subject to the direct influence of regional meteorological phenomena (weather). No sunlight or rain falls into a sheltered cavity and no regional wind passes through it. For meteorological effects to penetrate far into a cave, there must be a physical agent which itself penetrates the cave. There are two such agents: air and water. Since air is the medium in which meteorological changes take place, this agent will be examined first.

There are three main processes, which are at least theoretically distinct, that cause the movement of air in caves: the chimney effect, the drainage effect and the entrainment effect.

The chimney effect may be observed in caves which have

at least two entrances, each at a different altitude. The term 'entrance' is not restricted here to cavities accessible to man or animals; any cavity (eg. a network of fine cracks and fissures) which will allow the passage of quantities of air is an entrance in this context. Air movement occurs when the air temperature outside the cave differs from that within the cave, producing a difference in air densities. The situation is best examined by a calculation of the pressure balance (or, more usually, imbalance) at the lower entrance. Wigley and Brown (1976) give the following general equations (see also fig.2):

At the lower entrance (L), the external pressure (normal atmospheric pressure) is -

$$p_{ext} = p_0 + \overline{p}_{ext} \cdot g \cdot h$$

Where p_0 is the external pressure at the altitude of the upper entrance (U), g is the gravitational acceleration, h is the altitudinal difference between the two entrances, and $\overline{p}_{ext}$ is the mean density of the air column of height h above L outside the cave.

At L, the internal pressure (pressure exerted by the air within the cave system) is -

$$p_{int} = p_0 + \overline{p}_{int} \cdot g \cdot h$$

Where $\overline{p}_{int}$ is the mean density of the air column of height h above L inside the cave.

When $p_{ext} \neq p_{int}$ air movement will occur. The densities of the two air columns are the active factors and these are a function of virtual temperature (infra). If the air, both inside and outside the cave, were perfectly dry, the imbalance would simply be given by air temperature. When it is warmer outside than inside (summer), the interior air will be denser and air movement will be in at U and out at L. The inverse is true when

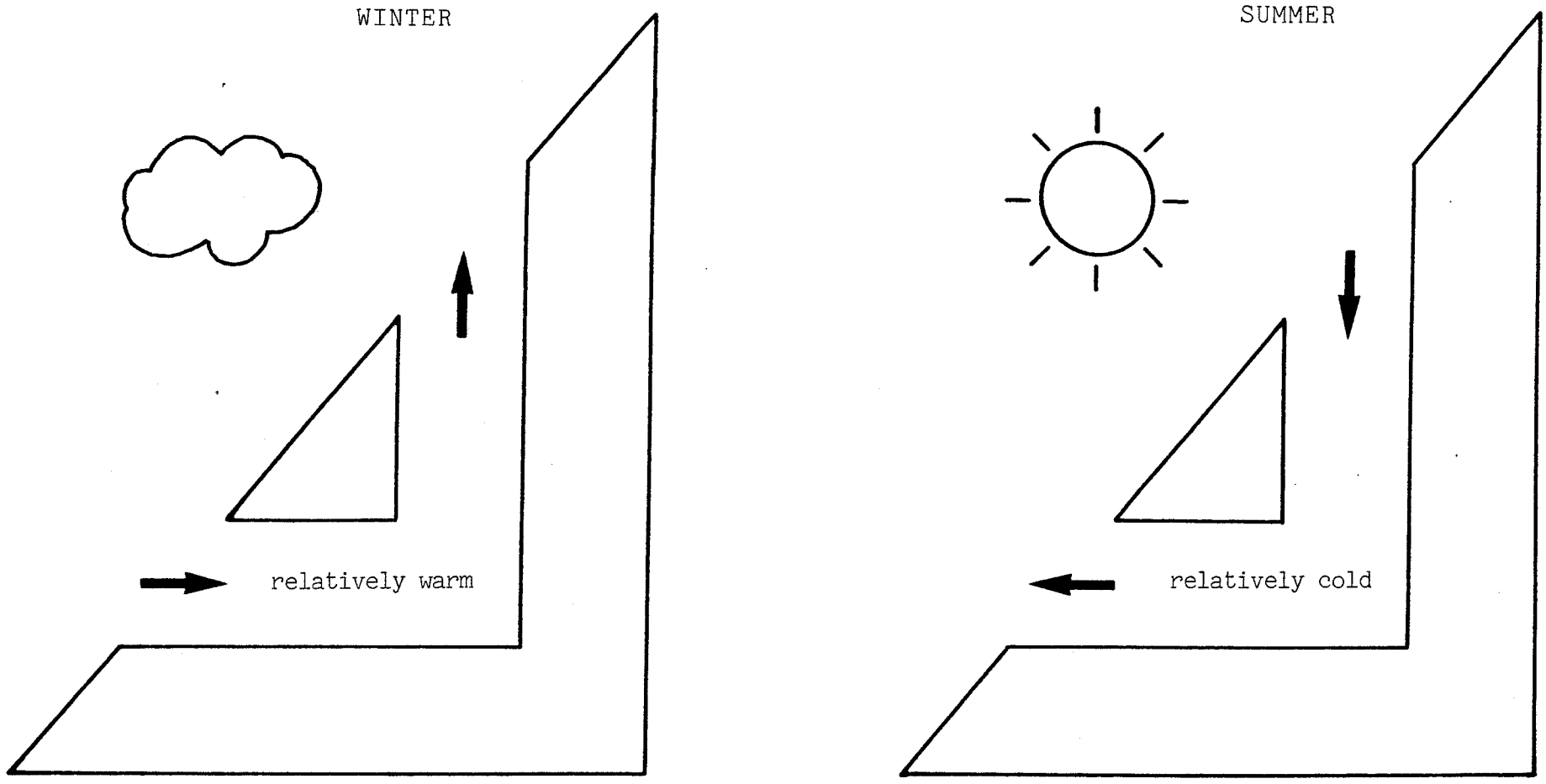


Fig.2 Theoretical Chimney Effect

it is warmer inside the cave (winter). During autumn and spring, the imbalance may oscillate producing more or less frequent reversals of air movement. In summer, dropping external temperatures at night will certainly slow the air movement and may even cause a temporary reversal. Bakalowicz (1972) has recorded diurnal variation in August at the Pinargözü Cave (Turkey) which follows a remarkably smooth rhythm, dropping to 1.3m/sec around midnight (16°C external temperature) and rising to 2.7m/sec at midday (30°C external temperature).

Air density is also a function of humidity and that is why virtual temperature is the exact parameter of interest. The virtual temperature is the temperature that dry air must have in order to reach the same density as the moist air in the situation under examination. Moist air is less dense than dry air (the apparent molecular weight of dry air is 28.964, whilst the molecular weight of water vapour is 18.015), so that rising humidity increases the virtual temperature. For example, saturated air (100% relative humidity) at 0°C, 5°C, 10°C and 15°C has a virtual temperature of approximately 0.65°C, 5.85°C, 11.35°C and 17.00°C, respectively.

The speed of air movement is a function of the magnitude of the difference in exterior and interior virtual temperature and of the altitudinal difference between entrances (the driving force), and the resistance to movement set up within the cave. Resistance, which increases with the speed of air movement (negative feedback), depends critically upon detailed cave morphology and is extremely difficult to estimate in any real situation. It is this fact which means that this model of air movement must remain, for the present, on the theoretical level, although we may predict the direction of movement quite accurately.

Speeds of c.5m/sec (18km/h) may be considered as the maximum likely to be encountered in cave systems that are several kilometres long. In smaller caves, such as most British archaeological caves, speeds of c.1m/sec (3.6km/h) will rarely be exceeded. Much higher speeds (up to c.160km/h in generally 'windy' caves) have been recorded in narrow sections of otherwise wide passages in caves from around the world.

The drainage effect occurs in single entrance caves which have a significant slope (fig.3). In a down-sloping cave in winter, cold exterior air may drain into the cave along the floor, displacing air warmed by the surrounding rock which passes out of the cave near the roof. In summer, the cold air of the cave remains essentially static. The inverse pattern (movement in summer, static in winter) will occur in an up-sloping cave. Such air movement is usually much slower than that caused by the chimney effect, possibly because it involves flow in both directions along a single passage, with some mixing of air and a resulting blurring of the temperature differential. In a long, linear and regular cave with an appreciable slope and an entrance at each end, the chimney effect and the drainage effect might combine to give a year-long pattern of cold air draining down near the floor and warm air rising near the roof. Again, in a two entrance cave with a length of down-sloping passage immediately within the lower entrance, there may be an increase in cold air inflow in winter with a little return flow of warmer air near the roof of that passage. With an up-sloping upper entrance passage in summer, there may be a slight increase of inflow of warmer air near the roof with a weak return flow of cooler air near the floor.

The entrainment effect is caused by running water which

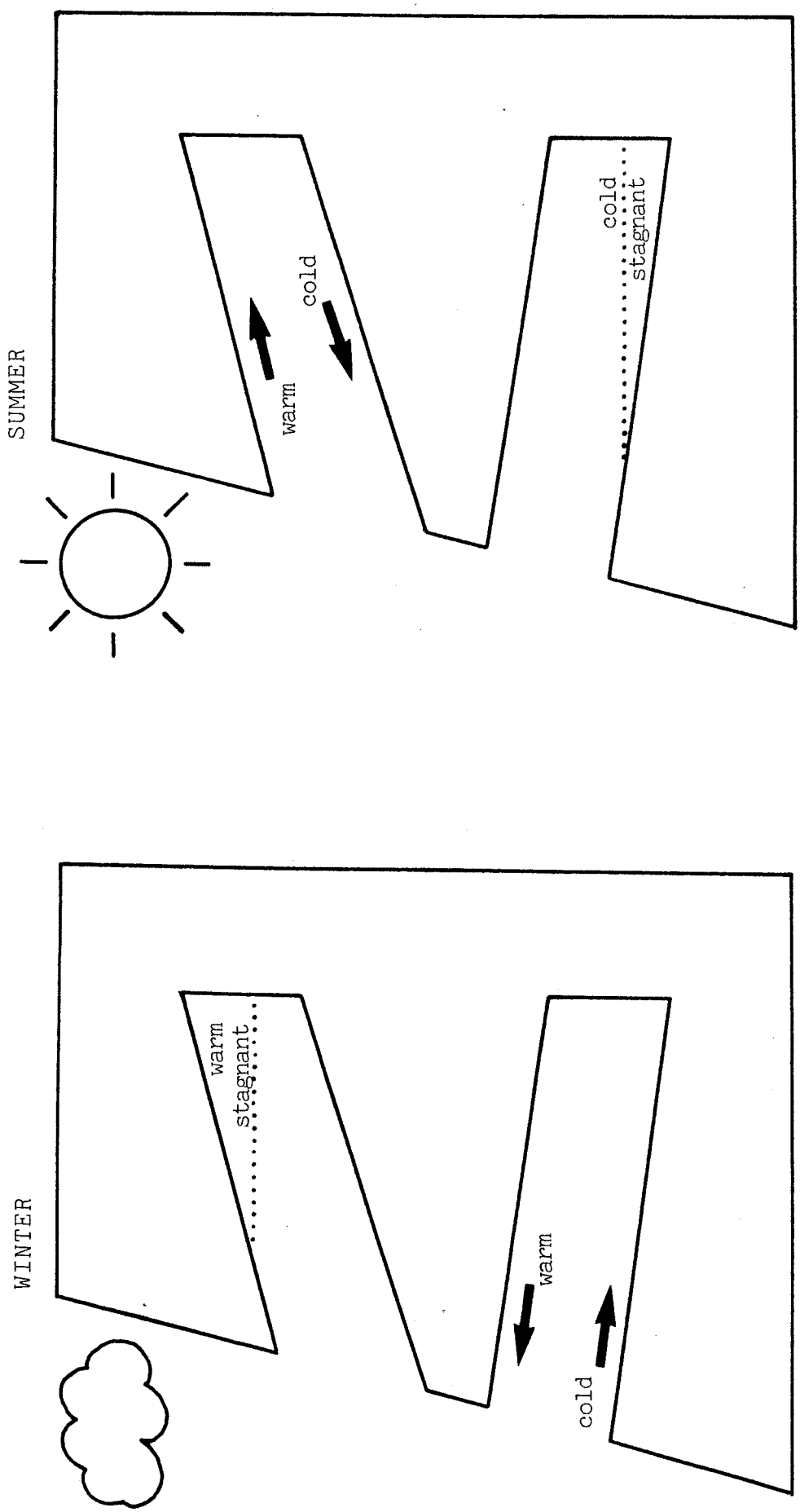


Fig.3 Theoretical Drainage Effect

passes on some of its momentum to air with which it is in contact (fig.4). In a single entrance cave (with respect to air not water movement), air will be pulled in and down near the stream surface and will be expelled near the cave roof. This will occur irrespective of season. Where the chimney effect and/or the (air) drainage effect are also operative, the entrainment effect will strengthen or weaken the prevailing movement, according to season. The entrainment effect is a function of the speed of water flow and the nature (roughness and area of the interface compared with the volume of air above it) of the air/water interface. The effect is particularly marked near waterfalls.

Other factors (eg. external atmospheric pressure fluctuations, resonance of cave air or rapid flooding) may cause movement of air, but the effects are only important in rather large cave systems. Air movement may be appreciable especially in a small multiple entrance cave if one entrance faces more or less directly into the regional prevailing wind.

Air movement is often a function of virtual temperature, which is itself a function of temperature and humidity. Since these last two parameters inside caves are in turn a function of air movement, it is clear that none of these constitutes a truly independent variable. However, it will be convenient to ignore the speed of air movement for the moment, because its primary effect is the depth to which exterior conditions modify the cave atmosphere; it has only a minor effect upon the form of the temperature and humidity curves (infra).

Wigley and Brown (1971) provide a numerical model for the distribution of temperature and humidity in caves. They make certain assumptions and approximations in order to simplify the calculations. Heat transfer within the rock walls is ignored

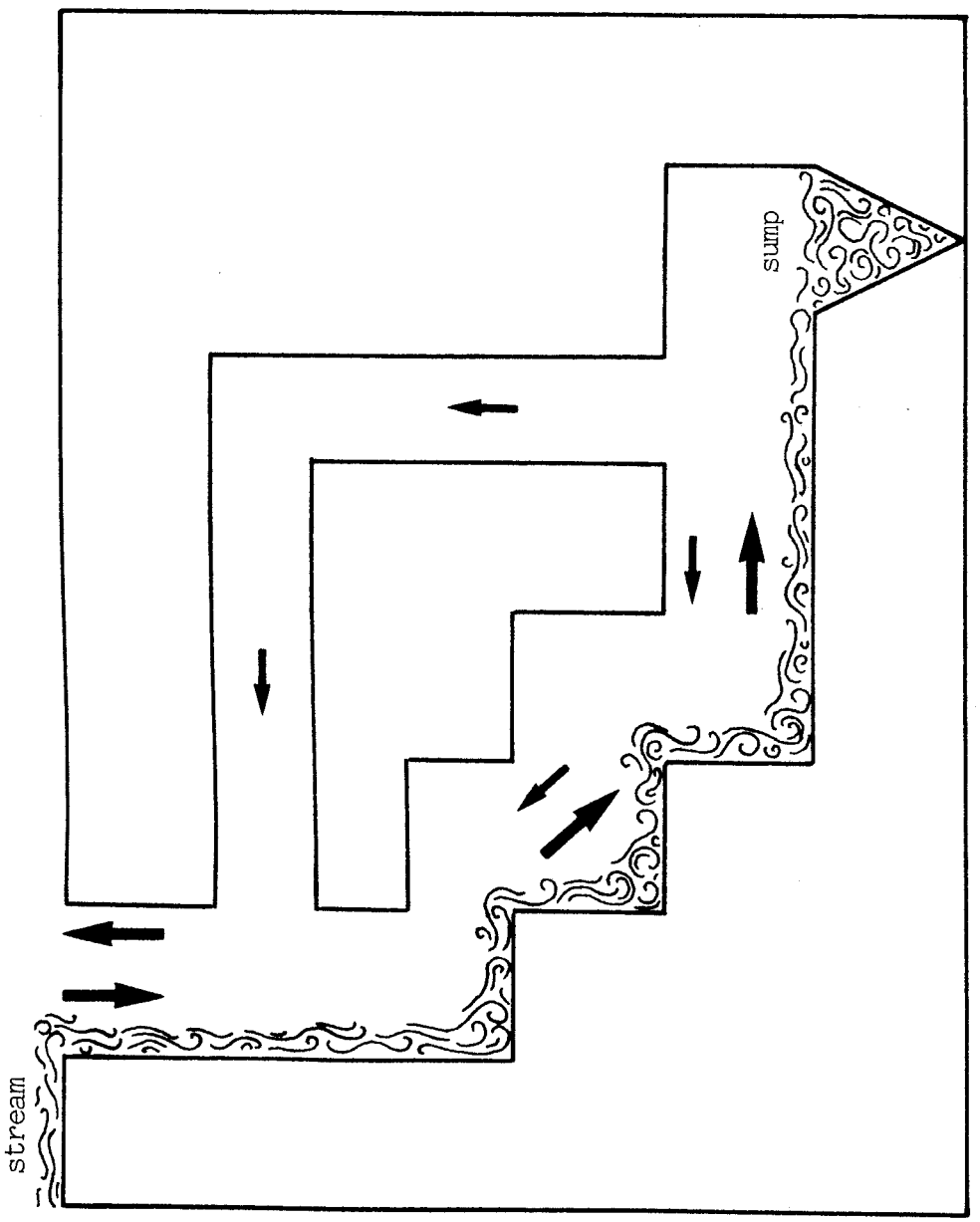


Fig.4 Theoretical Entrainment Effect

and only the conditions at the air/rock interface are considered. Since we are interested in meteorological changes along a passage, some variables (air flow speed, temperature and humidity) may be averaged over the cross-section of the passage. Air flow is assumed to be turbulent, an assumption consistent with the great majority of observed cases, and fully developed for heat, momentum and mass transfer. Viscous energy losses within the air are negligible and certain flow parameters (eg. the Lewis number) may be approximated in order to remove terms from the equations.

There is a further set of assumptions which involve departures from probable real conditions:

- The passage is horizontal, allowing pressure effects to be ignored. However, a more realistic situation can be modelled by including terms for adiabatic lapse rates.

- The passage is of constant radius and therefore the air speed is constant. Wigley and Brown (1971) give alternative, but approximate, equations if variation in passage diameter is to be considered.

- The rock wall is moist at all points. Wigley and Brown (1971) give a method for allowing for departures from this assumption.

- The temperature of the rock wall is constant and equal to the asymptotic temperature (the theoretically constant temperature deep within the cave). Here, the (untrue) assumption is made that the rock may cool or warm the cave air without itself changing temperature. It is only justified since the thermal diffusivity of the rock is much less than that of air. Thermal diffusivity is a measure of the facility with which a material will undergo temperature change. With figures for dry air, the low density (0.00129g/cm^3 at 1atm) ensures a low

volumetric heat capacity (0.000312cal/cm^3) which makes up for the high heat insulation potential (thermal conductivity 0.0000573cal/cm/sec/°C) and gives a relatively high thermal diffusivity (0.1839cm^2/sec). For an 'average' limestone, although the thermal conductivity is high (0.006cal/cm/sec/°C) the density is also high (2.7g/cm^3), producing a high volumetric heat capacity (0.567cal/cm^3) and a resulting thermal diffusivity (0.0106cm^2/sec) that is very much lower than that for dry air. As the air becomes more humid, the thermal diffusivity drops but only by a negligible amount. That air is a bad thermal conductor does not matter because it is a fluid that does not have to rely upon conduction for heat transfer away from the rock/air interface. Thus, we may allow rapid heat exchange at the interface without significant change in the temperature of the rock away from this interface. It follows from this approximation that the water vapour pressure at the moist interface will also be constant and equal to the saturation water vapour pressure at the temperature of the rock. On p.57 we will return to the possibility that rock temperature may be roughly constant at a point but may vary as a gradient against depth into the cave.

Given the above assumptions and approximations, Wigley and Brown (1971) propose the following equations for temperature and humidity distribution:

$$\overline{T} = T_a + (T_0 - T_a)e^{-x} + \frac{L}{c}(q_0 - q_a)xe^{-x} \qquad (1)$$

$$\overline{q} = q_a + (q_0 - q_a)e^{-x} \qquad (2)$$

Where $\overline{T}$ = air temperature (K) at a distance (x relaxation lengths, infra) into the cave,

T_a = asymtotic temperature (K) = wall temperature (K),

T_0 = air temperature (K) at the entrance,

e = base of natural logarithms (2.718),

x = distance (relaxation length, infra) into cave,

L = latent heat (vaporisation 2.501×10^6 J/kg; sublimation 2.834×10^6 J/kg),

c = specific heat or air at constant pressure (1005J/kg/K),

q_0 = specific humidity of air at entrance (g/kg),

q_a = saturation specific humidity at T_a (g/kg),

$\overline{q}$ = specific humidity (g/kg) at a distance (x relaxation lengths, infra) into cave.

Specific humidity values are best calculated from saturation vapour pressure / temperature tables (eg. Houghton 1977), using the formula:

$$q = (0.622 p_v) / (p_a - 0.378 p_v)$$

Where p_v = saturation vapour pressure (mb),

p_a = atmospheric pressure (1013.25mb).

Latent heat is not a constant but a function of temperature. However, the value for 273.15K (0°C) may be used here, since there is always less than 3% deviation from this value over the temperature range of interest (-10°C to 30°C). Nevertheless, good linear approximations are available if required, eg.:

$$L_{vap} = (2.5 \times 10^6) - (T - 273.15)(2366)$$

As long as the $\overline{T}/x$ graph is plotted at small increments, the $\overline{T}$ value for the preceding plotted point may be used as T to calculate approximate L, in order to avoid the coupling. The correction does not change the shape of the graph to any significant extent.

Specific heat of air is not a constant but a function of humidity. However, the value for dry air may be used here since any likely correction for humidity will not be greater than 0.5%,

as can be seen from the following approximation:

$$c_{humid} = c_{dry}(1 + [(0.9q) / (1 - q)])$$

Consider the general equation:

$$\overline{z} = z_a + (z_0 - z_a)e^{-x} \qquad (3)$$

Where z = T or q.

Equation (3) defines the exponential decline of z with distance into the cave, x. Equation (2) is of this form and Equation (1) reduces to this form when $q_0 = q_a$ (which is equivalent to stating that the external dew-point temperature is equal to T_a). The righthand term in Equation (1) allows for the effects of water phase changes (evaporation, condensation, freezing, melting, sublimation) and the resulting latent heat transfer. When $q_0 > q_a$, moisture will condense (or freeze out) from the cave air, liberating heat which warms the air. When $q_0 < q_a$, evaporation (or sublimation) occurs which cools the air. When air temperatures are below freezing, the value of L is changed and tables of p_v over ice rather than over water must be used to calculate q.

When condensation is occurring ($q_0 > q_a$), the cave air will not reach saturation (100% relative humidity), that is, $\overline{T}$ will always be greater than the dew-point temperature of the cave air. However, when evaporation occurs ($q_0 < q_a$), the air will eventually reach saturation, thus preventing further evaporation (and evaporative cooling of the air). Once the air has reached saturation, Equation (1) is no longer applicable. The saturation point is found by comparing values of $\overline{T}$ and $\overline{q}$ until the dew-point temperature is reached. Beyond this point, a new set of equations (Wigley & Brown 1971) is required:

$$A = [q_a L] / [cR_v(T_a)^2] \quad (4)$$

$$B = L / [R_v(T_a)^2] \quad (5)$$

$$x - x_s = (1 + A)\ln[(T_s - T_a) / (\overline{T} - T_a)] + AB(T_s - \overline{T}) \quad (6)$$

Where R_v = gas constant for water vapour (461.52J/kg/K),

x_s = the value of x (relaxation lengths, infra) at saturation,

T_s = the value of $\overline{T}$ (K) at saturation,

and the other symbols as previously defined.

Relative humidity remains at 100% at points beyond x_s.

The above equations may be solved (using at least a microprocessor) by substituting values of T_a, T_0 and q_0. Representative values for external temperature and humidity are comparatively easy to measure or to derive through pre- or postdiction. However, the choice of a value for T_a is far more difficult.

The parameter T_a is strictly the temperature of the rock wall. The fact that T_a is here equal to the asymptotic (deep cave) temperature is only incidental, since this follows automatically from the assumption of constant T_a. The only way to be certain that a measured deep cave temperature is indeed the asymptotic temperature is to demonstrate that there is no significant annual variation. Alternatively, if an accurate mean of what little variation does occur deep within the cave can be calculated, this mean can be used as the theoretical asymptotic temperature. Because of these difficulties, published values of asymptotic temperature must be treated with care in case they were incorrectly derived over too short a period. Nevertheless, a rather interesting pattern emerges which gives a clue to the

factors involved in true T_a. In caves that lie quite close to the surface (within c.30m) for most of their length, the asymptotic temperature closely approximates the mean annual external temperature (cf. Andrieux 1973). In caves that penetrate far into a rock massif, the asymptotic temperature tends to depend upon external conditions in an inverse manner. Myers (1962) reports that cave temperatures in English caves (which are not particularly deep) are often slightly below the annual external mean. As a generalisation, this is true of 'temperate climate' caves that are not overlain by several hundred metres of rock. However, in 'cold climates', where permafrost or glacier ice is present at or near the surface, deep cave temperatures are usually above the annual external mean (cf. Ford et al. 1976). These observations suggest that, close to the surface, the asymptotic temperature is linked to external temperature through a directly proportional relationship. Far from the surface, the relationship, although less regular, does seem to be the inverse of the amount and persistence of ice near the surface. The ice temperature itself cannot be the active factor or the relationship would be direct not inverse.

Pure ice has a thermal conductivity (0.0052cal/cm/sec/°C) only marginally lower than that of limestone, but some partially saturated fine sediments, with ice immobilising soil moisture and atmosphere, are slightly better insulators (cf. Linell & Tedrow 1981, fig.4.7, pp.34-5), whilst snow is very much better (thermal conductivity 0.0012cal/cm/sec/°C at 0°C and density 0.42g/cm^3; cf.Weller & Schwerdtfeger 1971). However, this thermal insulation cannot be the most important factor either, since we would then expect asymptotic temperatures to approximate the annual external mean.

Before exploring the possibility of mass transfer of heat other than by air, the discussion must be widened from the concept of spatially constant T_a to include any likely variation in rock temperature with depth. We must first decide whether there is any 'baseline' temperature gradient that might be expected in the rock itself. The low thermal diffusivity of rock means that heat will not be transferred very far and that gradients will decline rapidly away from a heat source. Nevertheless, near the surface, external temperatures must set up a gradient. Andrieux states:

> En supposant qu'un massif karstique constitue un milieu homogène (ce qui est inexact comme nous le verrons plus loin) et sachant que l'écoulement de la chaleur se fait par thermoconduction, il est possible de calculer la temperature à différentes profondeurs et à des temps variables.
> Dans ces conditions et compte tenu de la diffusivité thermique du calcaire, on calcule que les amplitudes thermiques journalières les plus importantes (30°C par exemple) ne sont plus ressenties au-delà de 7 à 8 mètres de profondeur. En ce qui concerne les amplitudes thermiques annuelles, même en considérant des valeurs de 40°C (ce qui représente une amplitude très importante) elles ne sont plus ressenties au-delà de 25 à 30 mètres de profondeur. (1976:130)
>
> [Trans.: Supposing that a karst massif constitutes a homogeneous environment (which is inexact as we shall see presently) and knowing that heat transfer occurs by thermal conduction, it is possible to calculate the temperature at different depths and at various times. Under these conditions and taking into account the thermal diffusivity of the limestone, it is calculated that the greatest daily temperature amplitudes (30°C for example) are no longer perceptible beyond a depth of 7-8m. Annual temperature amplitudes, even considering values of 40°C (which represents a very great range), are no longer perceptible beyond a depth of 25-30m.]

Note that 'depth' in this context refers to thickness of overburden, not to distance into a cave passage. The above statement is most surprising and it appears to contradict all that is known about heat transfer in rock. For instance, Coutard et al. (1974) give the results of laboratory measurements on various limestones which suggest that the penetration of fluctuations in temperature is an order of magnitude lower than Andrieux's estimates. Dresch (cited in Coutard et al. 1974)

states that, in limestone, a daily range of 30°C is no longer perceptible after 1.33m, a figure much more in keeping with the thermal diffusivity of most limestones. There is no obvious way to reconcile this disagreement, since the most conductive limestones (or even quartzites) would not give the figures suggested by Andrieux if thermal diffusivity is the only factor involved. Even if Andrieux's 'homogeneous environment' does not mean 'homogeneous rock', the inclusion of a regular pattern of voids, filled with air, water and sediment, would most probably reduce the thermal diffusivity even further. Note that, if Andrieux is in fact right, the neglect of heat transfer within the rock itself in the Wigley and Brown model (cf. p.51) would introduce unacceptable error.

Despite the apparent uncertainty, it is clear that, at the point where annual amplitude drops to zero, the temperature will be approximately the same as the mean annual external ground temperature. This is certainly the case, for example, in sediments subject to permafrost (cf. National Academy of Sciences 1974). Note that mean annual ground temperature is not necessarily the same as that of the air. It has been reported as being 1.5-9.3°C warmer than for air in permafrost areas around the world (Grave 1967) and there are theoretical reasons for believing that this might be a general tendency under most climates (cf. continuous geothermal heat loss, infra). The mean annual ground temperature will therefore serve as a starting point for any other temperature gradient occurring within the rock.

Beyond the superficial rock affected by external conditions, the geothermal gradient becomes more a function of the transfer of heat from the Earth's interior towards the surface, where heat is being lost at a planetary average rate of

62.79 ± 6.3 J/m²/sec (Lee 1963). The exact gradient at a given location depends upon the strength and depth of the heat source (mostly radioactive decay in igneous rocks) and the thermal diffusivity of the overburden. Thus, in volcanic areas and younger orogenic belts, the gradient may be as steep as 1°C/10m, whilst in 'old' areas of the crust covered by deep sedimentary sequences, the gradient may be as low as 1°C/100m. The 'average' gradient may be taken as an increase of 1°C for every 30m of overburden (Terzaghi 1952).

Combining the above information, it seems reasonable to assume a 'baseline' temperature gradient of c.1°C/30m, with a starting point, equal to the mean annual external ground temperature, situated probably within 2-3m of the surface but certainly not deeper than 30m.

This gradient helps to explain some of the observations of 'asymptotic' temperature already noted. Its recognition also shows that the very concept of 'asymptotic' temperature can never be more than an approximation. Caves remaining within c.30m of the surface over most of their length will tend to have wall temperatures close to the mean annual ground value. Even if the gradient starts actually at the surface, the maximum rise in temperature (1°C) may be offset because evaporative cooling is to be expected for more than half of the year (Wigley & Brown 1976; infra) and because of water infiltration (infra). The gradient also explains the high wall temperatures in some deep caves. However, there is still no explanation for the low temperatures in other deep caves nor for the apparent link with surface ice.

It is, of course, obvious that water, as well as air, can produce mass transfer of heat within a cave. If a true

stream flows into a cave, the water temperature will normally have had sufficient time to reach, or approximate, equilibrium with the external atmospheric and ground conditions. The stream may therefore be either relatively warm or cold with respect to the cave air. Although stream temperature will vary as a function of general external temperature, response to change will be slow and temperature amplitude will be greatly reduced, so that a stream will usually have a moderating effect on cave climate. The heat transfer between two highly mobile fluids (air and water) is difficult to calculate in any real situation. The larger the stream in proportion to the cavity, the more internal conditions will deviate, all along the passages, from those predicted by the numerical model discussed here. Naturally, there will also be slow heat transfer between water and rock. More empirical data is needed before more than a very rough qualitative appreciation of the effects of an inflowing stream can be achieved.

Water may also reach a cave passage by infiltration along microcavities in the rock. Far below the surface, the geothermal gradient will be such that the rock would normally be comparatively warm. Water reaching this area by infiltration will be comparatively cold, especially when it is derived from ice or snow melt. Note that the quantities of water passing through microcavities may exceed those concentrated into coherent streams by several orders of magnitude, depending upon the structure of the rock and the catchment. As long as water circulation is slow (the thermal diffusivity of water, 0.014 cm^2/sec, is only slightly above that of limestone), heat transfer will approximate to equilibrium conditions, with the rock temperature (and the rock surface temperature, T_a, in the

cave) being depressed below the value predicted by the geothermal gradient alone. If water circulation is fast (through more important fissures), quantities of cold water may even reach the cave passage itself, resulting in direct heat transfer between the water and the air, cooling the latter (as in the case of a true cold stream). Andrieux (1974) gives temperatures for an area deep within the Grotte de Niaux (Ariège) which has over 300m of rock overburden. Temperatures at this point should be constant (asymptotic) and somewhere in the range 20-28°C if the geothermal gradient is all that is involved. However, measured mean temperature is 12.5°C for the air and 10.8°C for the rock, with fluctuations of up to 3°C amplitude. Measurement of drip and sheet flow water temperature at this point shows a mean value almost identical to that of the rock and a very strong positive correlation between the two, but the amplitude of water temperature variation is slightly higher. This last point shows that the rock temperature is partly a function of water temperature, since heating or cooling of the rock is slower than changes in water temperature brought about by changing rates of water circulation. The air temperature also shows a strong positive correlation with that of the rock and of the water. That an abrupt change in water temperature, with a slower change in rock temperature, may still correlate with quite a rapid change in air temperature, shows the effect of direct water/air heat transfer. In general, the larger the fissures connecting a cave with the surface, the greater will be the variation in water circulation speed and thus the amplitude of temperature variation in the water as it reaches the cave. Correlated variation in rock and air temperature will result.

The effect of water, especially infiltration water but

often stream water as well, is to reduce, sometimes drastically, the warming effect of the geothermal gradient in deep caves (cf. Andrieux 1976). This explains the low 'temperate climate' asymptotic temperatures that have often been observed. It is now also clear how the presence of ice near the surface will more or less effectively reduce water circulation and allow a closer approximation to the normal geothermal gradient.

No generally applicable model can be proposed to include the effects of heat transfer by water, since circulation patterns vary so much between individual caves or even between sites within a single cave. However, the following guiding principles may be useful:

(1) In near-surface caves, the mean annual external temperature (ground temperature, if known) may be used to approximate T_a. T_a may be lower, but never below 0°C, if there is much water infiltration. If the cave passage lies only a few metres from the surface, seasonal or even diurnal variation in T_a is to be expected, due to external fluctuations in both temperature and fluid water availability.

(2) In caves with greater rock overburden, under external climatic conditions which do not cause extensive and prolonged freezing near the surface ('temperate' caves), the geothermal gradient will be lowered as a function of water infiltration volume and circulation speed, which depend upon the external precipitation/runoff ratio and the morphology of the microcavities in the rock.

(3) In caves far from the surface under external climatic conditions which cause long periods of ice or snow persistence near the surface, the geothermal gradient will be decreasingly affected as water circulation is reduced. This means that,

even in areas where the mean annual external temperature is below zero, a zone where temperature never drops below zero will eventually be reached if the cave continues far enough below the surface.

(4) In all real situations, T_a will never be constant, either in terms of distance into the cave or of time. The only way to produce realistic values for T_a is to set up an empirical model for a given cave based upon extensive measurement data. Nevertheless, a constant T_a model is useful as a generalisation, as long as the above points are kept in mind.

Returning to the constant T_a model, Wigley and Brown (1971, 1976) give representative curves for both temperature and relative humidity (reproduced here in modified form as figs. 5 - 7). They assume a relatively well ventilated cave and external conditions of a 'temperate' kind. Obviously, the curves only apply to these exact conditions, but many of the interesting features of cave meteorology are well illustrated. Wigley and Brown distinguish three types of profile:

(1) The summer profile (fig. 5) occurs when both the external temperature (25°C) and the dewpoint (21°C) are greater than the asymptotic temperature (10°C). As damp air moves into the cave, the release of latent heat as water condenses on the cold walls slows down the normal cooling trend, so that the decline is less abrupt than under the simple exponential case. Note that, in the example figured, condensation at first causes a drop in relative humidity of the air itself; the air may be drier than outside for a considerable distance into the cave and, even deep within the system, relative humidity may never reach 100%. At the particular values of external conditions shown in fig. 5 , latent heat release at first outweighs direct cooling

Fig.5

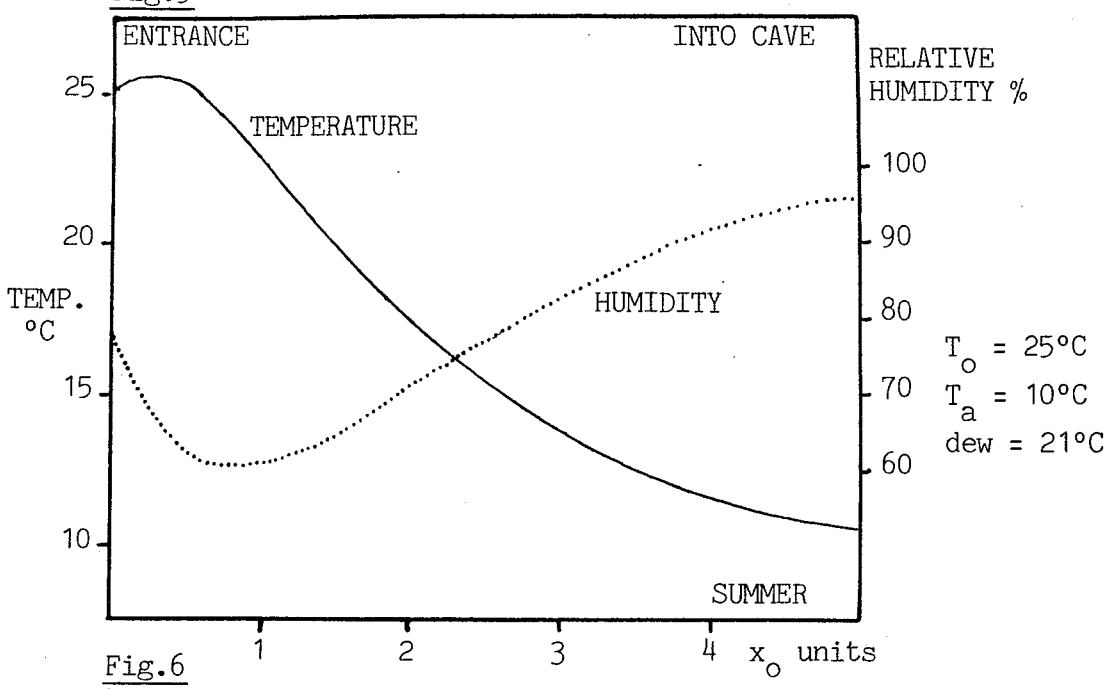


Fig.6

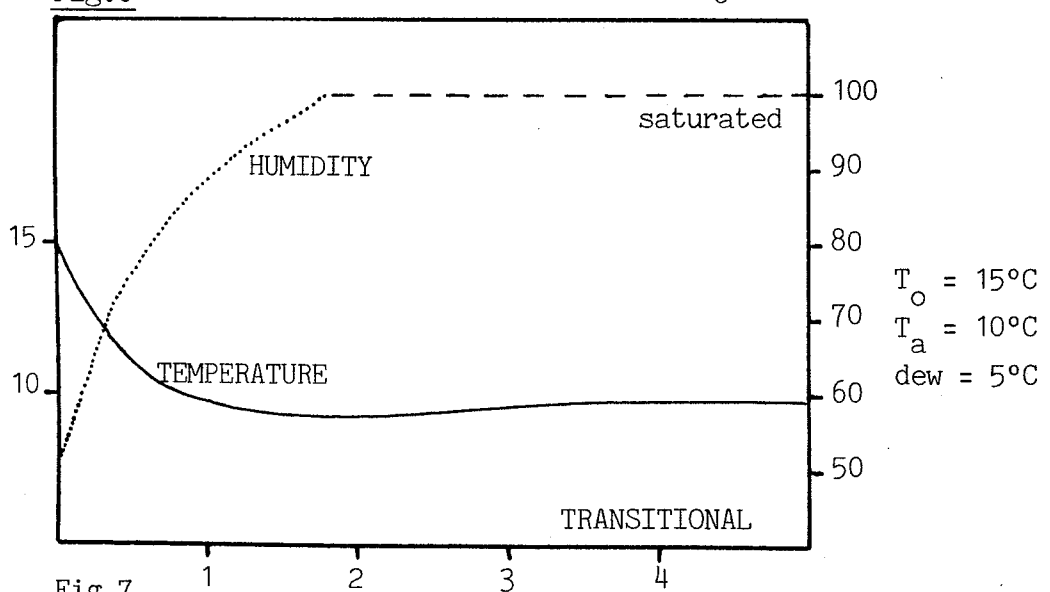
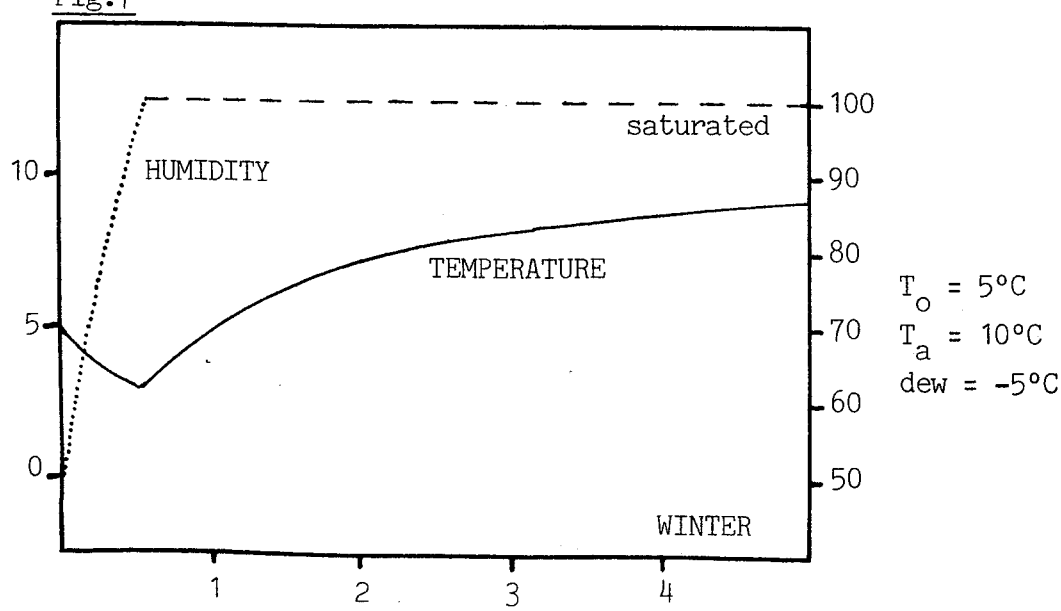


Fig.7



Figs.5-7 Cave Meteorology - Temperate Case

from contact with the cave walls, so that a zone with air temperatures slightly higher than outside is created near the entrance. If external relative humidity (and thus dewpoint) were not so high, this zone would not form, nor would there be an initial drop in relative humidity.

(2) The transitional profile (fig. 6) occurs when the asymptotic temperature (10°C) is intermediate between the external temperature (15°C) and the dewpoint (5°C). Temperature drops faster than exponentially as the air is cooled both by contact with the walls and by convertion of heat as moisture evaporates. This cooling overshoots the asymptotic temperature to produce a cold zone some distance within the cave. The lower the external relative humidity (and thus the dewpoint), the more pronounced will be this cold zone. Note that relative humidity rises quite rapidly within the cave due to uptake of water by evaporation. The turning point in the temperature curve is defined by the point at which relative humidity reaches 100% (evaporation ceases). The air is then warmed a little to reach the asymptotic temperature by contact with the walls.

(3) The winter profile (fig. 7) occurs when both the external temperature (5°C) and the dewpoint (-5°C) are lower than the asymptotic temperature (10°C). The rock walls are now much warmer than the air, so that an exponential rise in temperature towards the asymptotic value might be expected. However, the dry air first picks up water by evaporation and it is therefore cooled rather rapidly, producing a cold zone near the entrance. With more continental climates, where the difference between the asymptotic and external winter temperatures becomes increasingly large, this discrete cold zone becomes less important. Nevertheless, Bögli (1980) reports that evaporative drying of

the cave walls at Hölloch (Switzerland) can reach up to 2km into the system by the end of winter.

The linear scale (distance into the cave) must now be calculated. In figs. 5 - 7 , units, termed x_0 units by Wigley and Brown, are shown along the abscissa; these units are also known as relaxation lengths. They refer to exponential or pseudo-exponential curves and are akin to the concept of 'half life' in radioactive decay. They characterise not only the rate of change in temperature and humidity but also the exponential decay of the amplitude of external temperature and humidity variations in the cave air (not the rock) at increasingly deeper points within the cave. Wigley and Brown (1971, 1976) have derived a relationship between relaxation length (x_0), passage diameter (D metres) and cross-sectional mean air flow speed (V m/sec), given turbulent flow (the most common situation):

$$x_0 = 100D^{1.2}V^{0.2} \qquad (7)$$

It is clear that passage diameter is more important than air flow speed. To illustrate this point, consider a theoretical cave which has a winter profile as shown in fig. 7 . We wish to know how far in the coldest point would be; this point would also be approximately the point at which relative humidity reached 100%. Fig. 7 shows that this occurs at c.0.6 relaxation lengths. Our cave passage is of moderate dimensions, with a diameter of 5m. The coldest point would be 250m within the cave with an air speed of 0.080 m/sec (0.288 km/h). The air speed would need to be 2.572 m/sec (9.259 km/h), that is some 32 times greater, to push the coldest point to double the distance, 500m inside the cave. However, with an air speed of 0.080 m/sec, we would only need to increase the passage diameter from 5m to 8.909m to push the

coldest point to 500m. Apart from showing the relative effects of passage diameter and air flow speed, these calculations also show the general scale of the system: although attenuated, external conditions may still penetrate far into a cave.

In Equation (7) above, it is clear that V, and thus x_0, must take positive values. This is equivalent to stating that air movement must be inwards with respect to the entrance situated at x = 0 in the model. In a two entrance cave with seasonal reversal of air movement, the initial point must be shifted between the two entrances. Andrieux (1976) has suggested that, with deep caves where significant geothermal heat is involved and very cold external climate, air movement would not reverse in summer but would continue to follow the winter pattern. However, minor routes for air movement (fissures) would become increasingly important and such a constant flow direction could be easily 'short-circuited', especially near the surface. It is also clear that the model only applies to situations in which the flow can be safely assumed to be dominantly uni-directional. If significant return flow is present (as is necessarily true of all one entrance caves), this model cannot be used.

Figs. 5 - 7 represent generally 'temperate' external conditions. Before examining the possible implications of the meteorological theory developed so far, it will be useful to introduce a theoretical example more in keeping with the cold conditions which would have existed during much of the Pleistocene. Figs. 8 - 10 represent conditions similar to those thought to have prevailed at c.10,500 BP. in southern England. Summer maximum is set at 13°C, winter minimum at -8°C and mean annual ground temperature at -1°C (permafrost conditions); relative humidity is 50% in the three cases, since there is no way as yet

Fig.8

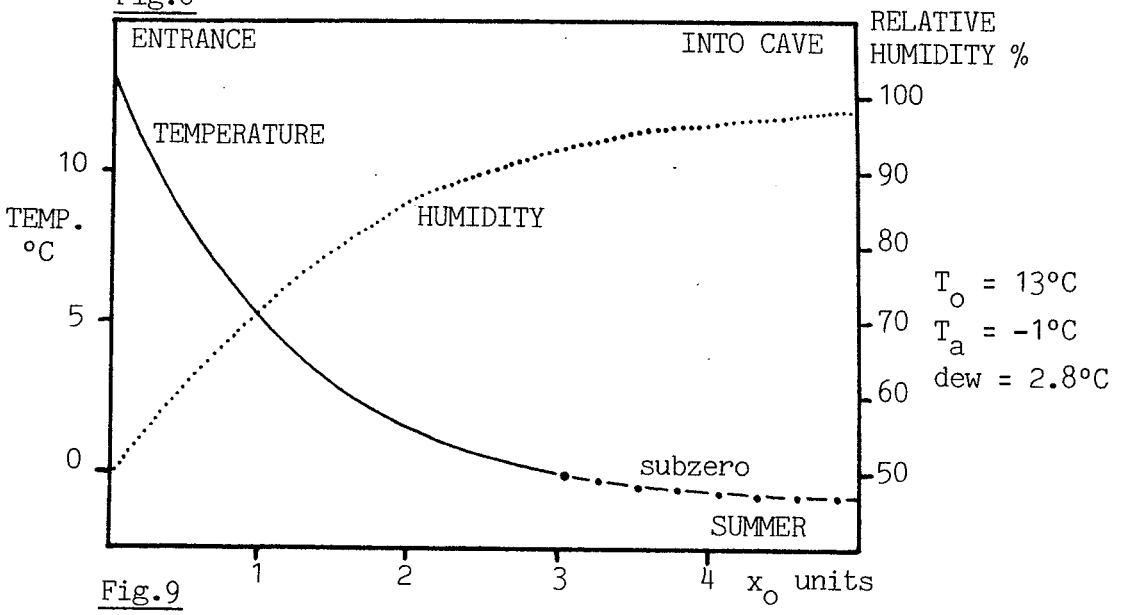


Fig.9

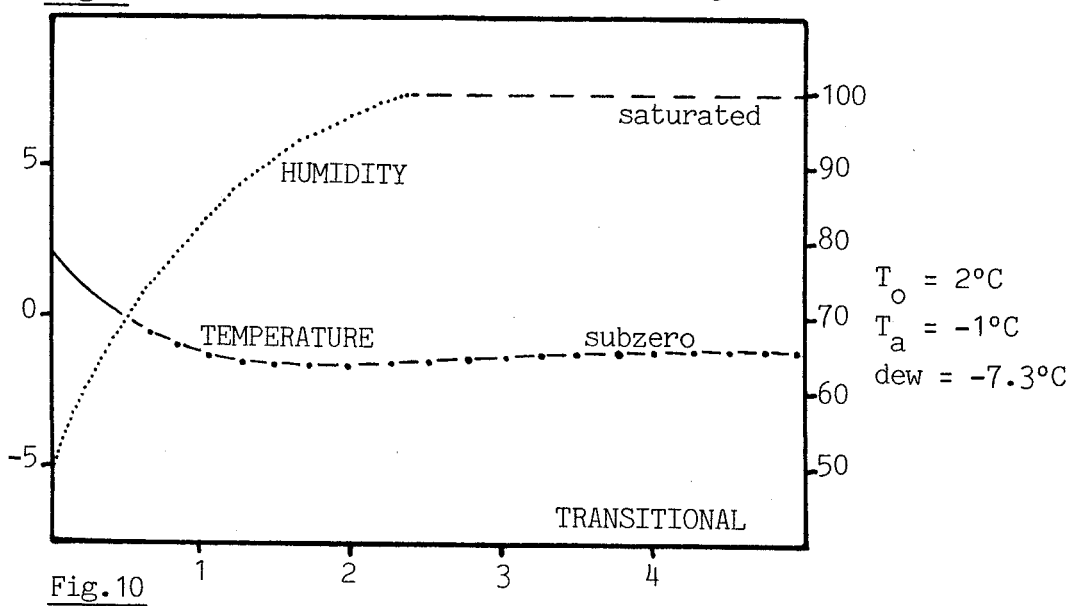
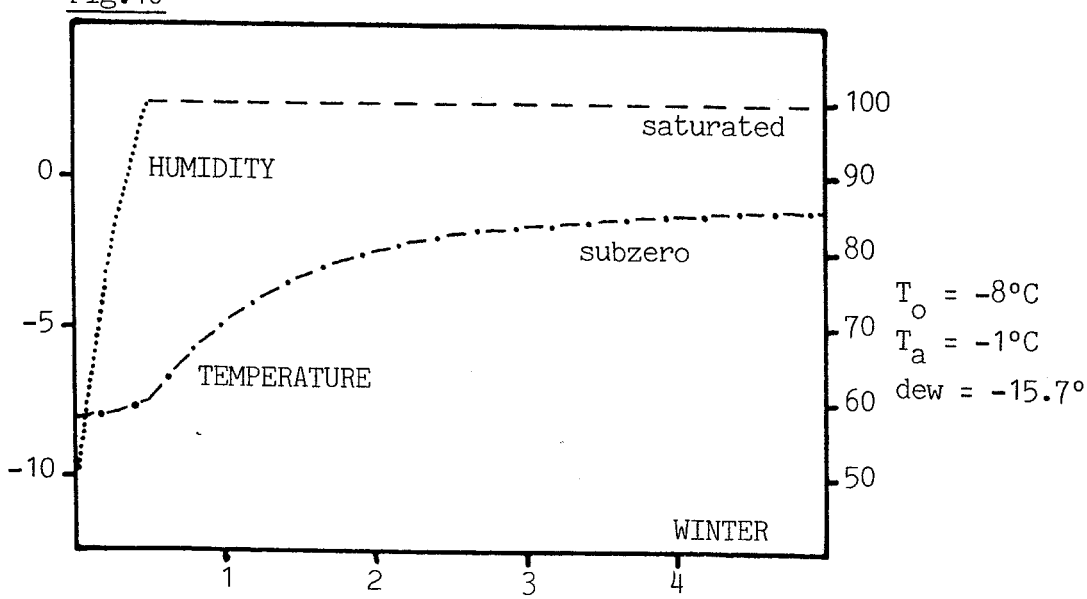


Fig.10



Figs.8-10 Cave Meteorology - Cold Climate Case

to estimate the probable values of this parameter. Note that, even if humidity were higher, there is no longer a warm zone near the entrance in summer. In winter, the cold zone is only represented by a slower rise in cave temperature; even at much lower humidities, the entrance area would not become significantly colder than the exterior. However, if T_a is allowed to be affected by heat transfer towards the surface, the winter cold zone may become more important since T_a will then be close to T_0 for the first few metres, allowing evaporation to cool the air to a temperature below that of the exterior. By c.30m below the surface T_a will certainly have risen to approximately the mean annual external level. Nevertheless, if the rock overburden (and thus the geothermal gradient) only increases slowly, the zone with sub-external temperatures may only start to decay after relative humidity has reached 100%. Theoretically, allowing T_a to vary in this way should also reinstate the warm zone, or at least a marked slowing of the temperature decline, during the summer. However, water circulation may reduce this effect to an insignificant level.

The meteorological model can be used to discuss the likelihood of ice formation in caves. It is assumed that air temperature is the important factor here, although it is difficult to see how ice would persist on rock that was much above the freezing point. Fig. 8 shows sub-zero air temperatures in summer at 3 relaxation lengths. This situation is a direct consequence of the assumption of constant T_a (infra). In winter (fig. 10), water cannot penetrate the cave since it will be frozen at sub-zero external temperatures and there can be little or no condensation of moisture from the air. Only superficial sublimation deposits ('hoar frost') will form.

It is the transitional case (fig. 9) which is of greatest interest. External temperatures are above zero and water is therefore available to form ice if it can reach the cold zone. Note that, if T_a is allowed to vary, there will still be a zone with sub-zero wall temperature and air temperature will be even lower due to the evaporation effect. The lower the external relative humidity, the more likely will be the formation of ice, given the availability of water. Once ice has formed extensively, rock/ice and air/ice interfaces must be considered and heat transfer may become rather more complicated. However, since ice (at least in the dense crystalline form) has a thermal conductivity and a thermal diffusivity (0.0112 cm^2/sec) similar to those of limestone, these complications may perhaps be ignored. The effect of porous sublimation ice is unpredictable. Ice will only disappear if summer conditions can introduce enough heat energy via air or water. Therefore, ice is more likely to form and persist nearer the lower entrance of a two entrance cave (cf. the seasonal reversal of air flow).

Most of the larger perennial ice caves in the world probably owe much to the processes suggested here; they are termed "dynamic" ice caves by Bögli (1980) with reference to the continuous movement of air. The Eisriesenwelt ('the World of the Ice Giants', Austria) has ice up to 1300m from the lower entrance (Hauser 1926). Castleguard Cave (Alberta, Canada) has ice in winter from near the lower entrance to 450m (according to Ford 1975) or even 1000m (c.3 relaxation lengths according to Wigley and Brown 1976) into the cave. This ice is destroyed by glacial meltwater in summer. The main upper entrance, c.9.6km away below a glacier, is permanently blocked by ice, but there must be other entrances since the air flow is of the normal, seasonally

reversing type. Water left in pools during the summer melt freezes nearer the lower entrance at the onset of winter. The regime very near the lower entrance (for the first 150m beyond daylight) is rather different. Extremely cold winter air (and a low mean annual external air temperature of 0.3°C) cools the rock so that it remains perennially below zero. During the short summer, cool but damp air flows out of the cave, giving rise to sublimation ice deposits. August air temperature, measured at a depth of only 18m and within sight of daylight, was -2.2°C, whilst the external air temperature was 23.3°C (Ford 1975). In winter, this ice is removed, again by sublimation, as dry cold air moves into the cave.

It can therefore be seen that great lengths of cave passage may be subject to a seasonal freeze-thaw cycle. However, mechanically significant short term cycles, which are dependent upon oscillations in external temperature and humidity, will occur only nearer entrances, since the amplitude of meteorological fluctuations in the air declines exponentially into the cave. At Castleguard, diurnal oscillations in April die out after 300m with an air speed of 0.67 m/sec (2.4 km/h) in a passage of c.1.5x4 m (Ford 1975). In long caves, only external temperature fluctuations on a medium scale (several days), produced by cyclonic conditions, might be expected to induce air flow direction changes that might allow freeze-thaw deeper within an ordinary cave. There is a possibility that variation in water circulation near the boundaries of any perennial ice mass could cause localised freeze-thaw cycles.

In all three seasonal cases of the model (figs. 8 - 10), the deep cave temperature is theoretically below zero since T_a is assumed to be constant at -1°C. Given a real cave of only

moderate cross-section (eg. 4m), with at least a weak air flow velocity (eg. 0.1 m/sec), one relaxation length would represent a great length of passage (c.333m with D and V as suggested). The summer curve crosses the freezing point at three relaxation lengths (1000m). Allowing for the geothermal gradient, it seems extremely unlikely that T_a would still be at -1°C this far into the cave which, presumably, would be at some considerable depth below the surface. Similarly, the winter and transitional curves should become positive at depth, probably even before 1000m because there would be little or no interference with the geothermal gradient from circulating water. Thus, as at the Eisriesenwelt and Castleguard, there will be an inner limit to the ice zone if the cave is deep enough.

Unfortunately, no observational data are available concerning the extent to which cave sediments may be seasonally or perennially frozen. It is rapid thawing, caused by direct insolation and fast air temperature rises, that produces liquidised (mobile) sediments in the open (cf. gelifluction, p.329). The model for caves suggests that thawing would normally be very much slower, so that critical levels of interstitial water would not usually occur. Rapid thaw in caves is associated with the input of large quantities of water, producing subaqueous sedimentation or fast mass movement (cf. p.371).

There is another form of ice cave, referred to as "static" by Bögli (1980), which relies upon reduced air movement. Even in small single entrance caves which slope down from the surface, winter cold air may accumulate, freezing any infiltration water that arrives during the warmer seasons. In summer, a true temperature inversion is created and the cave can only be warmed slowly by conduction or by penetration of warm water.

The implications of the meteorological model for human (or animal) occupation of a cave are obvious. In hot climates, where cooler summer conditions are welcome, most caves offer considerable natural environmental benefits (cf. Kebara Cave, Israel; Legge 1972). In cold climates, only a few types of cave would appear naturally attractive. Single entrance, upward sloping caves, that are well insulated by rock, snow, ice or frozen sediment overburdens, would be acceptable in winter, although there might be problems with the smoke of any fire that was lit (infra). Downsloping single entrance caves will certainly be cold throughout the year, raising the interesting possibility of their use as summer 'fridges'. Two entrance caves do not as a rule offer significantly higher temperatures than the exterior. However, as with any cave, they do provide shelter from wind and direct precipitation, though air movement to be expected in such caves (either in or out of the entrance) may itself be a nuisance, especially in winter. At an air temperature of -8°C, a cave draught of only 1 m/sec (3.6 km/h) will produce an extra chill factor of c.200 kg/cal/m²/h, which in human terms is the equivalent of the energy expended in slow, cross-country skiing (U.S.Army & U.S.Air Force 1966). In summer, in addition to any chill factor (eg. 10°C, air flow at 1 m/sec, chill factor of c.150 kg/cal/m²/h), rock and sediment surfaces will be damp from condensation. A real advantage is to be expected in winter with the upper (out-draught) entrance of very deep, multiple entrance caves, where rapid air flow may transfer significant quantities of geothermal heat to the surface. Unfortunately, such entrances are usually only found at higher altitudes. To benefit from such a site, man would have to put up with the harsher conditions due to altitude during any external activities. However, true

hibernating animals, such as bears, would greatly appreciate such caves.

The choice of an inhospitable cave as a habitation site will depend upon two main factors. First, a sub-optimum cave may be chosen because other considerations (eg. protection from predators, proximity to an important resource, etc.) are considered more important. Second, if man wishes to occupy a cave, he does not have to put up with the natural environment within it. The use of fire seems to be the obvious answer but this is not as simple as it might at first appear. Lighting a fire just anywhere in a cave can often result in an alarming increase in local draughts. One may find oneself being roasted on one side, frozen on the other and choked with smoke. The trick is to place the hearth(s) correctly and, if necessary, further modify the air movement with windbreaks (including complete blocking of a passage to create a 'single entrance' cave). In a large cave, this may be difficult and the use of tents or other shelters may be called for. Since we might expect to be able to recover physical evidence of such structures in archaeological cave sites (cf. p.678), it would be interesting to see whether a site considered to be a 'base camp' (i.e. not just a casual, short term occupation) had an optimal organisation for a given season. Man's ability to change his environment even raises the interesting possibility that naturally attractive caves might have been actively avoided, in order to reduce competition with den carnivores which would certainly know of such caves.

The meteorological model developed so far applies to cave passages where heat transfer is mostly by convection or by conduction of geothermal energy towards the surface in the rock.

Once one moves into an area that receives heat by direct solar radiation, there may be a dramatic change in conditions. This difference is most marked at cave entrances and rockshelters that face more or less into the sun. Such sites may still be sheltered from the wind and direct precipitation and they have the considerable advantage that they will be warm during the day (solar radiation augmented by reflection and radiation from the rock) and quite warm at night (radiation from the rock heated during the day). The relatively low thermal conductivity and high volumetric heat capacity of limestone are therefore beneficial in this case, in as much as most of the solar energy absorbed is stored very near the rock surface and then steadily released during the hours lacking direct insolation. Andrieux and Bouvier (1967) have shown that, in certain caves and shelters in S.W.France, the warmest point is the geometric centre - one might say the focal point - of a cave entrance or a shelter's concavity. This point was also found to be the driest, although it also had the highest range of temperature and humidity. With a cave entrance, there would be the added problem of a draught.

The aspect of a cave or shelter has long been assumed to be an important factor in whether or not it was chosen for occupation. However, aspect is not the primary factor, since it is the quantity and duration of radiation that is important. It is not possible to calculate this from an ordinary map, as the specific horizon of a site is not apparent in two dimensions. It is no good a site facing into the sun if it is in shadow from topography to sunward for much of the time. The modern radiation budget at a site can be measured, using either a black bulb thermometer (cf. Legge 1972), which shows heat flux at a point in time, or by chemical methods (cf. Bouvier 1975), which provide

estimates of heat received during a given period. Rock temperatures are best measured with thermistors (cf. Goudie 1981:147). However, direct measurements may often be inapplicable to past colder conditions since, in present temporate zones, trees and other vegetation may shadow the site and otherwise complicate the energy budget. Empirical data must also be corrected for cloud cover. Bouvier (1975) has overcome these problems by constructing an accurate scale model (in polystyrene) of the area around Les Eyzies (Dordogne), in which there are so many archaeological cave and shelter sites. The model is illuminated by a lamp (the 'sun'), positioned according to the correct geometric relationships for magnetic bearing, latitude, season and time of day. Correction may even be made for past differences in the Earth's axial tilt. Although precision is lost due to the scale (1:10,000), valuable comparative information about sites in a small area can be recovered. There is no reason why a larger scale model for a single site and its specific horizon could not be constructed, so that variation in solar radiation at different points within the site could be compared. Indeed, the shelter of La Madeleine has already been accurately mapped by photogrammetry with such a study in mind.

The relative importance of heat reflection and radiation from the rock during the day and heat radiation during the night is difficult to calculate, since so many site-specific parameters are involved, even given an exact knowledge of the morphology and orientation of the rockface. A finer grained limestone will have a higher thermal conductivity than a coarse grained or porous limestone and will therefore store more heat (Coutard et al. 1974); this will be beneficial to the cave or shelter environment as long as penetration is not so great as to lose the extra heat

to mobile fluids (air or water) in fissures within the rock. Even values for the thermal response of a given type of limestone may depend to a considerable extent upon its surface state (intrinsic colour, weathering state, organic stains, etc.). Thus a dark stained rock with a lower albedo (the proportion of incident radiation immediately reflected from a surface), such as many of the present limestone surfaces in the Périgord Noir, might be expected to store more heat for release at night, whilst a lighter rock, as in the Périgord Blanc, would provide more heat during the day. Dark rocks have often been reported to be over 20°C warmer than the surrounding air after a few hours of insolation (cf. Souchez 1967; Thorn 1979). In colder climates (at constant latitude), with less dusty and less humid air, more solar radiation reaches the ground and the heat store effect is accentuated. Actual figures for the albedo of different limestones are not available. However, dark soils may give values as low as 0.03 and snow as high as 0.85, with concrete at c.0.25 probably representing material analogous to 'average' limestones. All that is certain as a general proposition is that some improvement of the thermal environment, including a rise in maximum and mean effective temperature (a combination of air temperature and radiant temperature which is a measure of human comfort) will result.

Unfortunately, very few actual site measurements are available at present. The Périgord project set up by Bouvier is continuing but no detailed results have yet been published. Legge (1972) reports a series of measurements of both air and radiant temperatures from the Asprochaliko Shelter (Epirus, Greece). This work was part of a project to compare this site with the nearby Kastritsa Cave. Although temperatures were only

measured for one day in summer and one day in winter at each site, the results will serve to illustrate the advantages of a sunny shelter compared with a more shady cave. Differences in mean temperature may only reflect topographic factors (the shelter is at a lower altitude and 6-7°C warmer). The most interesting result is the fact that, at the shelter, radiant temperature nearly always remained above air temperature, even during the winter day when there was almost constant cloud cover. At the cave, the opposite was true. Therefore, in the shelter, a human gains significant radiant heat whilst, in the cave, he loses it to the surroundings. Incidentally, further measurements taken during the winter day on the talus outside the cave showed that, if radiant heat loss occurs in the cave, such loss is over twice as severe on the slope away from the limestone heat store. The cave also had a significantly higher air temperature than the talus. These results support the prior suggestion (Higgs et al. 1967) that, during the Late Last Glacial when the two sites were occupied, transhumance was practised between the lowland shelter in winter and the upland cave in summer. Finally, it is interesting that the palaeolithic hearths at Asprochaliko had been placed against the rear wall of the shelter, thus making further use of the heat storage capacity of the limestone.

Given that sunny shelters and cave entrances have a significant thermal advantage in colder climates, we may ask whether aspect has any effect upon deeper cave climate. If air moved into a cave via a warm entrance, it seems unlikely that very much heat would be transferred to the interior. If the draught were weak, the heat input would be significant only in the first few metres. If air movement were strong, the residence time of the air in the sunny entrance would not suffice

to allow the air to warm up. However, the mean ground surface temperature is the starting point for the geothermal gradient. Thus, Andrieux (1976) reports deep temperatures of 12°C in the Grotte de Sainte-Catherine, developed from a south facing slope, but of only 10°C in the Gouffre de la Peyrère, developed from a north facing slope in the same massif in the Ariège. This effect is to be expected under most 'temperate' conditions but it is not so predictable in 'cold' climates. If the land surfaces were mostly rock, the effect would still operate. However, if much sediment were present, a reverse effect might operate, since the 'cold' slope might have insulating permafrost whilst, at least in summer, infiltration water from the 'warm' slope might penetrate the cave and lower the temperature. Snow would also have an insulating effect, but some solar radiation will still penetrate up to c.1m of snow cover to warm the underlying surface and mobilise water (cf. Geiger 1965). It is possible that the constructed pavements (Fr. "dallages") found in many archaeological shelters and cave entrances reflect the increased melt from ice and snow produced by better solar aspect (cf. p.678).

The effect of aspect on geomorphology and sedimentation in colder climates is well known. If rock or sediment can absorb significant solar radiation even when the air temperature is far below zero, it follows that there will be more frequent freeze-thaw cycles at ground level in more sunny sites (cf.Thorn 1979:218). Sedimentation rates will often be faster and, as long as erosion rates are not equally high (which would be strongly dependent upon microtopography), deeper sequences will usually accumulate.

Cave meteorology can be seen to be a complex function of many different parameters, including external climate, aspect,

ground temperatures, cave morphology and depth, rock overburden, and water circulation patterns. Most of these parameters interact in ways that can only be qualitatively assessed for the moment. Nevertheless, the principles set out in this section can often provide useful insight into the environment of individual sites, both from the sedimentological and archaeological points of view.

As a generalisation, we may say that man in Europe seems to have preferred shelter and cave entrance sites that provided naturally superior thermal environments. However, deeper and less sunny sites were occasionally selected, often for reasons that escape us at present, but sometimes for special functions (eg. cave art) or because it was relatively easy to change the natural environment. Cave meteorology is certainly important with respect to site selection, but so are functional considerations and cultural preference. For instance, at Cuzoul de Gramat (Lot) or the Abri du Martinet (Lot-et-Garonne)(cf. Boone 1976), is the general change of environment between the Pleistocene and the Holocene a sufficient explanation for, on the one hand, the 'magdalenian' choice of the deeper, more sheltered areas and, on the other, the 'sauveterrian' choice of the sunnier, more exposed areas? Such a question should be seen as a starting point, requiring detailed assessment of the probable seasonal variations of environment within different parts of a site during different periods and deductions concerning the function(s), season, duration and scale of the occupations.

2.3. Zonation

The discussion of meteorology presented above introduced

the subject of variation in cave environment. The range of variation falls into more or less obvious subdivisions, which may be expressed areally as 'zones'. Cave zones, although often defined by natural discontinuities, are in fact only useful to us if their recognition provides a framework within which we may examine particular topics of interest. Zones which are relatively warm or cold are of interest with respect to human occupation, but slight differences in relative humidity are of little importance in this context. However, the exact value of relative humidity near an entrance may define a boundary between an outer zone, where some airborne particles can still penetrate, and an inner zone, where the air is clean due to water droplet nucleation around dust particles and subsequent fall-out. Such zonation would be useful if we were considering pollen dispersal or the penetration of fine loessic material.

Meteorological variables are by no means the only factors which can serve to define useful zones. Zonation based upon one set of parameters will not necessarily coincide with zonation based upon another set. Furthermore, zones are not temporally constant and the time scales for changes may be very different as one changes perspective. Nevertheless, no zonation scheme ever reflects a totally autonomous system and more will be learnt about cave environment only if we keep a constant lookout for correlation between the factors which we have artificially isolated from the whole. Some of the principles of cave zonation are briefly set out below, with special attention to factors that might be expected to produce physical traces that will survive over long periods.

The physical setting (topography) of the cave and its surroundings may be mapped and zoned. At least the main features

of the local geomorphology will be of interest and, in some cases, extremely detailed surveying will be called for. Such work is often carried out at an early stage but there is no reason why increasingly complex information cannot be added as the need arises. Extraction of information from general, and often imprecise, published sources long after the completion of fieldwork may be a useful adjunct but it is a poor substitute for problem-oriented observation. Purely topographic information is the starting point for any survey; details of rock outcrops, prominent sediment bodies, soil types, etc., can be added as required. Accurate mapping is desirable but is less important than a good understanding of the general variability and spatial relationships within the landscape. Mapping in limestone areas, with frequent steep slopes and vertical cliffs, is rather difficult and the time needed to produce even a slight increase in accuracy can often be put to better use. An introduction to geomorphological mapping is provided by Crofts (1981).

As one approaches the cave itself, an increase in scale and accuracy will be necessary but the principles are similar to those employed in any wider survey. The basic framework is that used in all speleological surveys (cf. Ellis 1976); additional information, concerning such factors as bedrock lithology or the form of rock walls, may be included. Cave systems will very rarely be accessible in their entirety and, in some cases, the site of interest will be choked with sediment right up to the entrance. Knowledge of the surrounding area will sometimes allow inference about underground links to be made at a later stage after examination of the sediments.

Topographic zonation is often the starting point for schemes based upon further parameters (cf. cave meteorology).

However, topography alone serves to focus the attention on concepts of sediment production, storage and transport routes, given different physical and chemical processes.

The hydrological regime within a cave has already been introduced (section 1.4.), with the main subdivisions of inactive vadose ('fossil'), active vadose and phreatic zones. Although water circulation will have been more or less different at given times in the past, the topography of the area, which changes comparatively slowly, represents a limiting factor on the potential hydrological system. Areas of concentrated circulation within the 'inactive' vadose zone (vadose trickle, drip lines, sheet flow,etc.) may be mapped. Such circulation, associated with specific morphological features (fissures, avens, rock overhangs, etc.), may often recur or persist over long periods. The catchment of a flow, be it a stream or a drip line, is important. Tracing of cave waters may be achieved by analysis of transported sediment and water chemistry, or by dye-marking at possible influent localities (cf. Smith et al. 1976; Cryer & Trudgill 1981). Response rate is also crucial, since a system which has a high water storage capacity and is able to buffer external variation will have a very different environmental (and thus sedimentological) effect from one which allows rapid through-flow leading to alternating flooding and de-watering. Even drip water may be more or less periodic. The study of response rates may often lead to inference about the nature of inaccessible parts of the system. The circulation of water within a sequence of sediments would be of considerable interest in many studies, such as the movement of solutes (especially carbonates; cf. Chapter 13) or thermoluminescence dating (cf. Chapter 18). Unfortunately, practically nothing is known

about this last subject in the cave context.

A substantial body of information exists concerning biology in caves (cf. Vandel 1965; Jefferson 1976; Cubbon 1976), although much of this deals with evidence which will not survive for very long in the cave environment (cf. Chapter 14). Biological material can be divided into several classes: fossil material (considerably older than the deposits under study; eg. hydrocarbons in the limestone itself), allochthonous material (transported into the cave by water, air or animals; eg. pollen, humus, remains of prey, etc.), excreta (eg. bat guano), and the remains of organisms which died within the cave. Animals of this last class may be divided into several groups (cf. Jefferson 1976:368): troglobites (species which are restricted to subterranean environments), troglophiles (species which are able to complete their life cycle underground, but which are not restricted to caves), habitual trogloxenes (species which often use caves for part of the life cycle, but which cannot complete that cycle underground) and accidental trogloxenes (species which have entered the cave by accident; cf. pitfalls). Most of the species of interest in vertebrate palaeontology (den carnivores, bats, roosting birds, estivating amphibians, etc.) are habitual trogloxenes. Plants and lower life forms (fungi, actinomycetes, bacteria, etc.) could be classed in a similar fashion, although complications arise with groups that are capable of alternative life cycles.

It is obvious that the presence of an organism in a cave depends largely upon the nature of the existing food chains. The green plants, the photo-autotrophes, depend on sunlight for their energy. Many lower life forms are able to survive on complex organic matter, ultimately derived from outside the

cave; these are heterotrophic organisms. However, some cave bacteria are chem-autotrophes, gaining their energy from the oxidation or reduction of simple inorganic compounds. Although these bacteria could be seen as the starting point for a truly independent food chain, most organisms still derive their energy more or less directly from the exterior environment. All vertebrate habitual trogloxenes of interest to us must actually leave the cave to find their food, since potential prey (eg. insects) is never sufficiently common to support even small animals.

Caves can be divided into ecological zones or "domains" (Glennie 1965). The epigean domain is the normal land surface where external conditions prevail; this domain extends into the cave as far as the maximum penetration of direct sunlight. However, there is a rapid succession of micro-habitats, often represented by a decrease in the variety and complexity of higher green plants and in overall biomass as one proceeds inwards. There is some interaction between the light (and thermal) regime and the plants, since large growths may shadow or totally obscure the cave entrance. In real situations, the mapping of this domain may be extremely informative, since the plant communities are very sensitive to sedimentation rate. The pro-epigean domain is the zone of reflected light which, depending upon cave morphology and aspect, may extend for quite some distance underground. Here, only lower green plants survive; it is nevertheless remarkable how little light some algae require. The permanently dark zone, the hypogean domain, is necessarily devoid of green plants but the lower life forms, such as fungi and bacteria, are common although they represent a very low biomass. Where water infiltration is relatively

constant (eg. pools, active stalagmite, etc.), communities also involving lower animals (eg. various worms or crustaceans) may be important. The zone which contains the roots of plants, the endogean domain, may extend far beyond the surface occurrence of these plants; tree roots may be found hundreds of metres underground.

It is clear that a modern cave could be divided into a complex system of zones based upon such factors as topography, hydrology, meteorology, ecology, etc. These are the very factors which, together with features inherited from the source, govern the erosion, deposition and modification of sediments. The study of cave environment is very much an unrealised potential. In section 3.2. below, the concept of 'facies' will be introduced, a concept which promises to link environment and sediments in a far more realistic and informative manner than simple climatic determinism has ever done.

3. THE PROPERTIES OF SEDIMENTS

3.1. Particles and Sediments

In order to avoid the reductionist spiral that would leave us wondering what quarks are made of, it is clear that we must define a particle as any object that is of interest to us for a specific reason. This subjective definition allows us to use the term 'particle' for a pebble of sandstone or for a sand grain included in that pebble, depending upon our current needs. Note that there is now nothing whatsoever preventing us from calling a whole sand dune a 'particle' if we so desire.

A sediment is a collection of particles. In geology, sedimentary rocks are said to be formed from particles of pre-existing rocks (clastic material), from organic debris and from chemical precipitates. The factor which allows this grouping is the absence of extreme temperature/pressure at any time during the history of the sedimentary rocks. Geologists normally use the term 'sediment' for unconsolidated material. According to this definition, many sedimentary rocks are formed from the consolidation of sediment but others (the precipitates) are not. Yet most sedimentologists are quite happy to consider precipitates when discussing sedimentary environments. Thus Pettijohn writes:

> The sediment is generally, but not always, deposited from a fluid in which it was contained either in a state of suspension or solution. (1975:1)

Clearly, differing perspectives with regard to scale and

cohesion are stopping us from arriving at absolute definitions of 'particle' and 'sediment'. However, we are not so much interested in objects for their own sake as for the way in which they behave. A particle is an object which behaves as an individual when acted upon by a given process. A sediment is composed of a number of individual particles responding to a given process, whether or not the particles interact. These definitions place the stress firmly on the concept of 'process'. We can now see that, given the process of carbonate precipitation, a calcite crystal is a particle and the resulting (pure) sediment is stalagmite. Given the process of rapid flooding, limestone clasts, animal bones, stone artefacts and fragments of stalagmite are particles and the resulting sediment is a heterogeneous gravel. The various particles produced by sieving a weakly cemented sediment represent behaviour in response to the process of sieving. If we adopt this perspective, the concepts of 'particle' and 'sediment' are no longer indeterminate.

When confronted with some ancient deposit, we come up against a problem. Since we were not there to observe the process, our definitions still cannot be applied if purely linear reasoning is adopted. We need to make an initial 'guess' concerning the process, which will then provide a working definition of the relevant types of particle and sediment. The initial 'guess' is derived through the application of the Principle of Uniformitarianism (Actualism), originally formulated by Lyell (1830-1833). The Principle involves the assumption that natural laws are constant in time and space. Accordingly, if a geological observation can be explained by analogy with the results of a process which can be seen to be acting in the present, it is not necessary, or advisable, to invoke a hypothetical

process which has no modern counterpart. This is merely a special case of Occam's Razor. The rate and intensity of a process need not be uniform, nor need the combination of physical conditions (environment, geological 'ingredients', etc.) be identical at different periods (cf. Gould 1965).

Our 'guess', then, is made by comparing the various properties of the ancient deposit with those which we know to be the result of modern processes. This is an iterative method of reasoning, whether or not we are concious of the fact. An increasingly precise definition of the process leads to more detailed and directed observation of the properties of the deposit, which in turn loops back to tighten the definition of the process. This method is open to considerable error if the initial 'guess' was wrong. A stalagmite may be treated as a sediment when, in fact, it is disturbed and is therefore better dealt with first as a particle responding to the process of disturbance. A deposit may be treated as a sediment which has responded to a depositional process when, in fact, it is a large particle (eg. a limestone slab) which has responded to a post-depositional process (alteration and disaggregation). Such error can often be avoided if the normal rules of deductive reasoning are followed. First, the initial 'guess' should encompass all the possible (conceivable) starting points, not just the first one which comes to mind. Obviously, the more restricted one's experience of what is 'possible', the less likely it will be that the true process is included in the initial 'guess'. Second, at each stage of refinement, the current idea of the process(es) involved should be used to predict probable consequences with respect to the properties of the deposit. If applied rigorously, this procedure may sometimes compensate for lack of relevant experience by

showing that none of the suggested processes is likely to have given rise to the observed deposit. One may then concentrate upon widening one's experience in order to include the 'anomalous' properties and associations of properties. Specific processes are dealt with in detail in subsequent chapters. The remainder of this chapter will involve the discussion of the different properties of deposits and the types of information they carry.

Consider the smallest unit thought to be of relevance in a given context: a particle. This particle has composition, a combination of chemistry and internal organisation. There are many derivative properties that can be measured, including density, porosity, permeability, strength, solubility, magnetic properties, radioactivity or colour.

The particle is also finite, so that it will have mass (or weight), size, shape, roundness and surface texture. The last four properties cannot be measured on any absolute scale. Size may be variously defined as volume, length, width, ability to pass through a hole of given shape and dimensions, dimensions in an arbitrary or oriented two-dimensional projection, etc. Shape is either an approximation to an ideal form (eg. prismatic, spherical, heart-shaped, etc.) or some ratio of dimensional properties (length, width, thickness, maximum projected area, diameter of smallest circumscribed circle, etc.). Roundness is a property somewhere in between shape and surface texture. It is a comparison of the general shape and surface curvature with the curvature or angularity of individual 'corners' (i.e. it is a measure of the variability of curvature). Concepts of size, shape and roundness are discussed in detail by Griffiths (1967). Surface texture is the microrelief of a particle. The last three properties (shape, roundness and surface texture) are, to a certain

extent, dependent upon size. A small particle like a sand grain is best observed using some type of microscopy and the distinction between the three properties may become blurred. On a large particle like a quartzite cobble, more than three levels in the hierarchy may be useful (eg. the inclusion of macroscopic pits or striations as a phenomenon intermediate between roundness and surface texture).

The properties so far discussed may be considered to be intrinsic to the particle, but they are not necessarily 'original' or immutable. From the point of view of quaternary sedimentology, an original state can often be assumed. Limestone becomes interesting when it forms an outcrop and is available as a particle source; anything that happened prior to this may usually be seen as an original state, with no time dimension. A particle under study represents the balance between the original state and the environmental history of the particle. Shape, for instance, is a combination of original composition and the cumulative effects of physical and chemical alteration. Particles therefore have properties that have been developed at different rates and at different times. This may sometimes be expressed as a recognisable sequence. A hemispherical limestone cobble has a white crust on its curved surface but is closer to the assumed original state on its flat surface; we may deduce that it was broken after it was altered and rounded. An even more striking example of sequence is provided by recent studies of superimposed surface feature assemblages on quartz grains by scanning electron microscopy (cf. papers in Whalley 1978).

A particle has one further important property: its position in space at the time of its discovery. This is usually expressed as a general location and an orientation of one or more

of the particle's axes. If we were interested in the particle merely as an object, position would not be an intrinsic property. It is our 'behavioural' definition of a particle that requires that position be considered.

The properties of sediments are merely the sum (or perhaps the product) of the properties of its constituent particles. We might measure the orientation of a large number of pebbles in order to gain an idea of the average and the spread of orientation in the sediment. We may automate a measurement technique (eg. by sieving a sediment rather than measuring each particle) but the particles are still being treated essentially as individuals. However, it is often impractical to function at the level of the particle. The study of sediment colour or bulk chemistry will provide approximately the same average values as the study of these properties at the particle level, but the range in values will never be so great, no matter how many bulk samples are examined. A shift in perspective from 'particle' to 'sediment' involves a narrowing of the spread of observed values for a given property. Other properties (eg. acidity, plasticity or conductivity) we believe to be the result of interaction between particles but the interaction is so complex that we could never calculate the end result from a knowledge of the individual particles. The properties of sediments are therefore derivatives, despite the fact that it is often convenient to ignore this distinction.

Sedimentology is concerned with the way in which sediments behave. This is reflected in the division of sediment properties into two main groups: textural and structural.

The textural properties concern the size and form of particles and the 'average' way in which they are organised into

a framework geometry. Specific topics depend upon the kind of problem one is trying to solve. Particle size distributions, and the graphical and numerical methods for describing distributions, constitute a major field of study in their own right. The study of packing emphasises the distribution of particles and voids of different sizes and shapes, together with the nature of contacts between particles. This is obviously the basis for sediment porosity and permeability. If we think in terms of specific surface (the ratio between the surface area of a particle or void and its volume), we can characterise the nature of the sites available for either addition or removal of material by chemical or physical processes. Particle contacts and cohesion are a starting point for a study of the strength of a sediment, its resistance to shear, deformation or compaction. Essentially the same basic properties can be expressed in terms of fabric, in which particle orientation is emphasised. Macrofabric may be defined to include sand sized material or to apply only to material larger than sand, as required. Preferred plan orientation and preferred dip (imbrication) can be measured. Microfabric refers to the orientation properties of the fine matrix and (unless much sand is included) is usually measured indirectly. A preferred orientation in the matrix will cause some degree of anisotropy in properties such as permeability or conductivity. However, scanning electron microscopy has made even the coarser colloids directly visible. Analysis of thin sections, using a petrographic microscope at comparatively low magnification, permits a detailed appreciation of microfabric when a good range of particle sizes is present (including sand, but preferably without too many larger particles). The technique has been developed mainly for the study of true soils, but the

descriptive vocabulary (cf. Brewer 1964) may be useful for many sediments (cf. Goldberg 1979a). Other aspects of sediment texture, and even structure, may be covered at the same time, and the whole is often referred to as micromorphology. Fabric studies in general may also include an appreciation of the relationship between matrix and larger particles (the extent to which larger particles are floating in the matrix or are in contact one with another). Wherever possible, a distinction is made between primary textural properties, induced by initial sedimentary processes, and secondary properties, resulting from post-depositional modification.

The dividing line between texture and structure is not absolutely defined and varies slightly with individual usage. However, if texture is concerned with 'average' organisational properties of a sediment, structure refers more to specific geometrical properties and especially to ordered change in these properties. There is a tendency for texture to be perceived on a smaller scale than structure, although this is by no means an absolute requirement. The concept of bedding is central to structural studies. A bed is a body of sediment which has properties that lend it a consistent identity and which set it apart from adjacent beds with other properties. Bedding may be defined by changes in composition, size, shape, orientation or packing of particles. Beds are usually observed in section, but it is important to gain some idea of their three-dimensional geometry. The identification of bedding is subjective, since there is an underlying assumption that bedding is due to primary sedimentary processes. Thus, soil horizons or zones of cementation would not be used to define beds. However, in some cases, it may be difficult to differentiate between depositional

effects and post-depositional (subsurface) effects. If a bed has internal consistency, this implies that there was a certain regularity in the process(es) then active. In this sense, a bed is very similar to a 'sediment', as defined above. Similarly, a change in properties (beds) implies a change in process. The nature, degree and rate of change in properties necessary to define a change in bed depend primarily upon the processes thought to be responsible. Minor changes in properties, that probably represent only minor changes in process, will often be allowed to occur within defined beds. Nevertheless, even such minor changes are still important and should be reported. A bed may contain many thin subunits oblique to the major axes of the bed (cross-bedding) or parallel to these axes (laminations). Indeed, internal bedding may be just as complex (including factors of subunit geometry, composition and texture) as the features seen on the larger scale of defined beds. Relatively smooth changes may occur within a bed that are referred to as graded bedding; particle size is usually involved but other properties may also show this tendency, as in particle density grading. The terminology of bedding is discussed by Reineck and Singh (1980).

If all time were represented by sediment, we would expect gentle transitions between beds, since we know that most depositional processes change comparatively slowly. However, sharp boundaries are very common and they will nearly always represent erosive or non-depositional events. Boundaries are extremely important to stratigraphy and they are discussed in more detail in section 5.1. Here it will suffice to note that structure is a response not only to additive deposition but also to subtractive erosion. The upper boundary of a bed may sometimes be defined by an erosion surface that is very much younger than

the bed itself. At other times, an erosion surface may be the result of very short term fluctuation in the environment (eg. tides). An erosion surface has no composition, only form, but this form may still be diagnostic of a process.

There are many structures, known as bedding plane markings, that may be thought of as small scale erosional forms. Raindrop impact pits, desiccation cracks, rill marks and the traces left by a bouncing or rolling object are examples of such structures. Biogenic structures (ichnofossils), such as footprints, also occur. If the non-depositional process responsible for a structure persists over a significant length of time, the structure may no longer be expressed as a true bedding plane marking but may penetrate far into the underlying bed(s). Animal burrows, root channels and ice-wedges may be included here. Any depression in a surface may trap sediment so that such a structure may have apparent composition (eg. scour-and-fill structures that often contain well bedded sediment). When a process persists, it will rarely remain constant, so that the form of the structure will change. Accordingly, the fill of the structure will be more complex, as in burrows or ice-wedge casts that may show several stages of fill representing progressive shifts in morphology and exact location.

The minor erosional features mentioned above occur after the deposition of the bed in which they are developed. Strictly speaking, they should not enter into the definition of this bed, except as phenomena defining a boundary. However, if the general sedimentation pattern continues more or less unchanged, sediment above and below (and even within) an erosional feature may be considered to be part of the same bed. Thus an erosional surface (of any kind) does not automatically define a bed, but

may sometimes be termed syndepositional (or syngenetic, synsedimentary). Indeed, features such as cross-bedding are almost exclusively the result of repeated minor erosion that is an integral part of a generally stable environment.

An alternative way of viewing some bedding plane markings that do not involve actual removal of sediment is to consider them as deformation structures. A footprint is theoretically a deformation structure, although little more than the surface expression will be visible. An ice-wedge may deform pre-existing structure (eg. bedding) more noticeably. When considerable pressure is applied to a bed, by overlying water, ice or sediment, major deformation may occur. Note that high overburden pressure need not always result in clear structures but may merely produce uniform compaction. Variations in the properties of sediments with time (often involving water content) may cause deformation under normal gravitational force. The exact nature of the structures depends upon whether the materials involved have deformed regularly or have ruptured (cf. Reineck & Singh 1980; Collinson & Thompson 1982). Frost sorting, sometimes giving rise to patterned ground, and advanced bioturbation may be considered as types of deformation which may affect considerable depths of sediment. It is necessary to differentiate between penecontemporaneous (or metasedimentary) deformation structures, which form during or slightly after the deposition of the sediment, and deformation which occurs long after deposition. This is largely a matter of the importance which the observer attaches to the time scale involved. In practice, a penecontemporaneous deformation structure is said to affect one bed, or two beds and the boundary between them, whilst the surrounding beds are undeformed. Just as with syndepositional

erosional forms, the implication is that deformation is closely associated with the processes active in the environment at or about the time of deposition.

The term 'structure' may be applied to other features. Imbrication is strictly a textural property but, if bands or lenses of imbricated material are present, they are often referred to as imbrication structures. Pedogenic features should not be used to define bedding but a soil may be considered to be a structure or complex of structures. Diagenesis (cf. Chapter 16) may produce many different types of structure. Bones in anatomical ('articulated') order are structures. Groups of objects organised by man are structures.

This brings us back to the underlying distinction between composition, texture and structure. Composition will be inherited from the source(s), although modification of the composition may occur during transport and, of course, through post-depositional processes. Reineck and Singh suggest:

> The study of sedimentary textures. It includes granulometric study of sediments. Grain size, grain size parameters [summary statistics], shape and roundness, surface texture, and primary fabric are the important aspects of such a study. These features are controlled mainly by the medium and mode of transport, and to a lesser degree by the depositing medium. Thus, they provide more accurate information about the transporting medium and only limited information about depositional conditions. This leads to the fact that the study of primary sedimentary structures is most important in the reconstruction of ancient physical environments. (1980:8)

The "depositing medium" is often the same as the "medium of transport" (eg. water); the distinction in the above quote refers to the dynamic properties of the medium which are different during transport and deposition. These authors are primarily concerned with high energy regimes, in which sediment is efficiently transported by water and air. The distinction between transport and deposition may not be so crucial with some low energy processes, such as slow mass movement, where transport is over a

short distance and deposition is 'instantaneous' once transport has ceased. Nevertheless, it is clear that the sedimentologist's prime concern is the environment at the time and place of deposition, and that textural properties may often be uninformative or even misleading. Thus, Selley writes:

> Unfortunately this approach [attempting to deduce environment from texture] has largely proved unsatisfactory and the granulometric analysis of ancient sediments is a declining art, to the delight of laboratory technicians [and research students]. (1976:6)

Structure, then, is the key to depositional environment. This in no way means that we can ignore composition and texture. First, we often wish to know something about the source and transport process. Second, and more importantly, there is no difference between 'texture' and 'structure' in reality; these are merely two ways of looking at essentially the same basic information. The main difference between the information carried by a particle and that carried by a sediment lies in the significance of this information with respect to what we wish to know. We find the information from a sediment infinitely more trustworthy than that from a single particle when considering environment, despite the fact that we are usually observing the same properties. Similarly, texture is unhelpful when it represents only a point estimate. When the spatial variation of the same properties is considered, structure is perceived. With structure, we have arrived at a sufficiently detailed perception of the sediment to be able to rule out the possibility that what we are seeing is due to chance or inheritance. This explains the ambiguity in the status of some traditionally textural properties (eg. imbrication); they, like all structural properties, are not 'transportable' and they cannot therefore have been inherited.

At this point it will be interesting to classify the properties usually recorded from cave sediments. As a representative example, we may take the following list of properties recognised by Laville (cf. 1973, 1976a):

Bulk granulometry (texture)
Granulometry of elements >10mm (texture)
Thickness/length ratio of limestone clasts (composition + texture)
Recognition of 'fresh' fractures on clasts (composition + texture)
Recognition of fissured clasts (composition + texture)
Rounding of clasts (composition + texture)
Porosity of clasts (composition + texture)
Composition and shape of aggregates (composition + texture)
Recognition of stalactite fragments (composition)
Study of 2-10mm material (mostly composition, some texture)
Granulometry of material <2mm, before and after decalcification (composition + texture)
Calcimetry (composition)
Quartz grain surface texture (texture)
Measurement of ΔpH (composition + texture)

These are all compositional and textural properties measured, with considerable expenditure of time and effort, in the laboratory. Laville (1973) mentions a few structural properties noted on site, such as the thickness of beds seen in one or two sections or the rare instances of complex and highly visible structures (eg. laminations or convolutions). It is hard to escape the conclusion that cave sedimentologists have not yet capitalised upon the information available from the study of structure. Numerous examples of structures are given in subsequent chapters of the present work; these include phenomena which have already been observed in cave deposits or which are likely to be found once we begin to look for them.

During the above discussion, some readers may have raised a serious objection: as quaternary sedimentologists, our primary interest is not the environment of deposition within a cave but rather the regional environment ('climate'). The present author disagrees. First, without the contextual and stratigraphic information produced by detailed study of the

environment of deposition, all other quaternary disciplines are seriously crippled. Second, we have seen that an increasingly complex (structural) perception of the sediments leads to accurate inference concerning the environment of deposition. This does not mean that, by concentrating upon individual particles or low order textural properties, we can reconstruct exterior or regional environment with anything like the same accuracy, even allowing that we can somehow deduce the availability of sediment, in general or of particular types, in the landscape. Exterior sediments have their own structure; once they are moved into a cave this high level of information is destroyed and textural properties, inherited from a spatial and temporal range of environments, are mixed. The environment of deposition within a cave is of course itself influenced by the regional environment but, as was pointed out in Chapter 2, the relationship is extremely complex and no a priori model will yet serve to link the two unequivocally. Even the relationship between the environment of deposition and the overall micro-environment acting at that point within the cave is often obscure. Every effort should be made to extract climatic and general environmental information but the intrinsic limitations of the study material must be recognised. Our obligations to other specialists must not be neglected for the sake of some climatic wild-goose chase, especially since other disciplines (in particular the biological sciences) can often produce quite precise environmental data if the context is clear. In any case, the author is convinced that increasingly accurate understanding of the environment of deposition is by far the most promising way for sedimentologists to approach a better understanding of general environment.

So far, we have seen that structure, perceived at the level of individual sediment bodies, provides reliable information.

> In our review of the various environmental parameters and our evaluation of the criteria used to ascertain them, it is clear that the attributes used - textural, mineralogical, structural, and biologic - do not as a rule give unambiguous answers. (Pettijohn 1975:542)

This statement, whilst it supports the view expressed above, also suggests that even structure is not the ultimate property and that there may be yet more informative ways of perceiving sediments.

3.2. Patterns and Facies

Structure implies patterning and patterning implies the passage of time. Patterning may simply reflect sequence (a footprint was formed after the deposition of the deformed sediment). It may reflect a shift in the active process through time or a time differential effect of the process (graded bedding represents a change in the dynamics of the transporting medium or differential settling of the particles). It may reflect oscillations between deposition and erosion or non-deposition (most cross-bedding). Once time is involved, our perception shifts from a 'flat' equation between sediment and process to the more substantial implications of changing process. Most simple processes occur in any number of radically different environments. However, the way in which a process changes is usually diagnostic of a much smaller range of environments.

The International Stratigraphic Guide defines 'facies' as follows:

> Facies, in stratigraphy [and in sedimentology in general], can mean aspect, nature, or manifestation of character (usually reflecting conditions of origin) of rock [including unconsolidated sediment] strata or specific constituents of rock strata. It is also used as a substantive for a body of rock strata distinctive in aspect, nature or character. The general term 'facies' has been greatly overworked. Rock strata may show differences in facies of various sorts so that one may speak of lithofacies, biofacies, mineralogic facies, marine facies, volcanic facies, boreal facies, and so on. If the term is used, it is desirable to make clear the specific kind of facies to which reference is made. (Hedberg 1976:15)

The present section is concerned with sedimentary facies (usually abbreviated to 'facies' below), which may be defined as follows:

> A sedimentary environment is a part of the earth's surface which is physically, chemically, and biologically distinct from adjacent terrains. (Selley 1978:1)
>
> A sedimentary facies is the product of a depositional environment, a special kind of sedimentary environment. (ibid., p.2)

A more precise definition might replace "depositional" with "predominantly depositional", since surfaces of erosion and non-deposition which are due to processes that are intrinsic to the environment are also an integral part of facies.

Thus, just as 'sediment' is a response to 'process', so 'facies' is a response to 'sedimentary environment', and the same type of iterative reasoning is required to define a facies (cf. p.89). Note that the term 'facies' may often be applied (especially in the phrase 'lateral facies') to aspects of a sediment that are assumed to be the effects of sedimentary environment (or changing sedimentary environment), without specification of the latter.

The scale of a defined facies will depend upon the environmental importance which the observer attaches to its occurrence. A facies may include many beds of different types, many beds of similar types, only a few beds, a single bed or it may reduce to a single sediment type with its implication of process. There are even cases in which a component of a

sediment may be usefully defined as a facies. Sedimentary environments are usually perceived on quite a large scale and they reflect major discontinuities which are often the same as those that serve to define everyday usage (eg. a lake, a beach or a desert). A given environment is normally said to consist of a series of micro-environments, each of which is represented by one or more facies. Minor shifts in environment may be represented by subfacies. Note, however, that definitions of environments, micro-environments, facies and subfacies are not necessarily fixed; they depend entirely upon what a particular observer, at a particular time, wishes to know about the study material and upon the actualistic models available to him.

Changes in process represent changes in environment, which in turn are expressed by changes in facies. No environment is totally stable, so that there will be some deviation from the average conditions. If average conditions remain stable, there must be feedback in the system which prevents major deviation. The system as a whole will change abruptly only when the feedback potential (the 'inertia' of the system) is exceeded. Gradual change may occur, however, by a slow drift in average conditions. When greater time scales are considered, the slow drift itself is often seen to be subject to feedback. Patterns of sedimentation are built up of an infinite series of cycles, with different periodicities, superimposed one upon the other. Some cycles may take so long that we cannot perceive them within a given sequence of sediments. Some are so short that it is again difficult to perceive them. The mechanisms which cause cyclicity are often divided into two groups. Autocyclic mechanisms are inherent to a given environment (eg. stream channel migration or division). Allocyclic mechanisms are

external to a specific sedimentary environment (eg. climatic variation, sea-level change, subsidence and uplift). Some mechanisms which act over comparatively short time scales may be considered to be inherent or external, depending upon the aspect of the environment under study (eg. storms, floods, seasonality or variations in sediment supply). Of course, this division between autocyclic and allocyclic mechanisms is always subjective; even the mechanisms quoted are abstractions from a single, though infinitely complex, real system. It follows that the cycles themselves are abstractions, a fact which has important implications when we turn to the sequences of facies which represent changing sedimentary environment.

A simple cyclic change in sedimentary environment will result in repetition of the same sequence of facies, other things being equal. The sequence may be symmetrical (eg. ABCBA etc.), asymmetrical (eg. $A_1B_1CB_2A_2$ etc., where the thickness or degree of development of the facies with the subscript 1 is not equal to that of their counterparts with the subscript 2) or totally non-reversing (eg. ABCABCA etc.). Although the nomenclature is by no means universal or consistent, the first two types of repetition are often known as rhythms and the last as a cyclothem. In nature, rhythmic sedimentation tends to be a shorter term phenomenon, whilst most longer term repetition is cyclothemic. There are interesting links between kinds of natural equilibria and time.

However, in reality other things are rarely equal and, as was noted above, a sedimentary succession will usually show signs of many superimposed cycles of different period, so much so that it may even appear non-cyclic. Alternatively, there may be harmonics and cycles themselves may be ordered into

repetitive sequences, or megacycles (usually megacyclothems but rarely megarhythms). Since we cannot examine all variation at once, especially since much of this variation may be irrelevant to the subject in question, it is necessary to isolate an 'ideal' cycle and to relegate everything else that is happening, whether or not other cycles are present, to the status of random noise. It is therefore clear that a cycle may be recognised even when the observed succession deviates from the exact order of the ideal cycle. In the succession ACBCDADBACDABC, the cycle ABCD is still recognisable. The longer the succession and the more often a cycle is repeated, the easier it is to detect the cycle. The most commonly used method for detecting cyclicity is a search for the Markov property (cf. Krumbein & Dacey 1969; Schwarzacher 1969; Dacey & Krumbein 1970), a property which describes the tendency for a sequence to be significantly non-random. The sequence may be said to have 'imperfect memory' which allows it to 'select', more or less correctly, which facies 'ought' to come next. Other methods have also been used (cf. Anderson & Koopmans 1963; Davis & Cocke 1972). Most analyses start from the construction of a table showing the number of transitions from any one facies to any other facies. This in turn gives an upward-transition probability matrix (eg. out of the 4 cases in which facies A changes to another facies, there are 2 cases in which the change is to facies B, producing a probability of 0.5 for the configuration AB given A, etc.). Such a matrix is easy to comprehend and will be useful even if further numerical analysis is not required. More complex matrices may be constructed by asking such questions as 'given AB, what is the probability of the next facies being C?'. An example of the use of transition matrices and Markov chains is given in section 3.3.

Note that the recognition of cyclicity, or of any pattern for that matter, is never an end in itself. We always need to find out what a pattern means and this brings us firmly back to considerations of environments, processes and actualistic models.

Often a sequence of facies is not repeated at one locality through time but occurs repeatedly in different places. Such a sequence is usually called a motif (cf. Pettijohn 1975; Selley 1976).

Nearly a century ago, Johannes Walther formulated the concept of the "faciesbezirk" (facies area). The facies area is contiguous with a defined environment and comprises the areal occurrence of all those facies which the various micro-environments are producing. The term may be extended to apply to the group of facies itself, without specification of the exact areal configuration. Walther observed that, in the absence of major unconformities, a vertical sequence of deep marine sediments, for instance, is not likely to contain a facies representing a fluvial or desert environment. He therefore stated:

> ... primär sich nur solche Facies und Faciesbezirke geologisch überlagern können, die in der Gegenwart nebeneinander zu beobachten sind. (1894:979)
>
> [Trans.: ... primarily only such facies and facies areas [groups] may be geologically superimposed [stratified] as are to be observed side by side at the present time.]

It has sometimes been suggested in the later literature that this statement constitutes a 'law' - it does not, and Walther was well aware of the fact. The statement is based upon actualistic principles which assume unchanging physical laws but not uniformity of process or expression (cf. p.89). Walther knew that some ancient facies and environments will not have exact modern counterparts. Low altitude, mid-latitude glaciers no longer exist, but we may still interpret pleistocene

glacial deposits using models derived from modern glaciers, modified, in accordance with physical laws, to take into account the geographical effects.

If we observe sediments through time at a single location, we will see patterns of variation, some of which we will be able to resolve into cycles or motifs, that are meaningful in terms of temporally changing environment. If we observe sediments in space at a single time, we will see patterns of variation, due to spatially changing environment, that are the same as individual vertical sequences, which, when repeated, may be perceived as cycles or motifs. The way in which environment changes temporally is the same way in which it changes spatially. This is a very powerful proposition, since it allows us to use the readily accessible spatial distribution of modern environments and facies to interpret ancient vertical sequences. Similar modern environments will have similar distributions of facies, so that we may set up a facies model (cf. Walker 1976) which will be an ideal pattern representing the common elements of those real environments which we choose to perceive as belonging to a useful class (cf. Reading 1978; Blatt et al. 1980; Reineck & Singh 1980; Hallam 1981). Unlike a defined cycle or spatially recurrent sequence (motif), a facies model automatically includes environmental connotations. Our job is therefore to find the facies model which best fits (explains) the observed cycle or motif.

> The next step is the study of lateral and vertical relationships of various units and the determination of their geometry. Then these can be compared with models of comparable environments based on information from the study of present-day environments. The special features of the ancient sediments should be discussed, and an attempt should be made to find out under which special conditions such features were formed. (Reineck & Singh 1980:7; emphasis added)

Increased understanding is only achieved by examination

of the deviation of the real case from the current model. This is akin to the "argument by anomaly" (Gould 1980), which is proving so useful to modern archaeology, and the rejection of the "reinforcement syndrome" (Watkins 1971), which for so long has bedevilled quaternary chronostratigraphy.

The enormous advances in sedimentary geology that have been achieved since the 30s owe a great deal to the universal acceptance of the importance of facies models. Cave sedimentology lags far behind other fields, mainly because of the lack of studies of modern cave environments (cf. Chapter 2), but there is every reason to believe that effective facies models can be set up for cave sediments. As a model represents the ideal, extracted from a large number of observed situations, it is very difficult for individual researchers to make much headway. Facies analysis in caves will only become a powerful tool if enough researchers co-operate to establish the requisite data base.

Lais (1941) noted the presence in many caves of cycles, alternating between coarse sedimentation on the one hand and fine sedimentation and carbonate mobilisation on the other. Using a combination of theory, general observations of modern environments and palaeontological data, he proposed that these two groupings represented colder and warmer conditions, respectively. Similar models, with increasingly detailed equations between the composition and texture of the sediments and the nature of the prevailing climate, as well as an increasing neglect of any actualistic data base, have formed the basis for most cave research ever since Lais's work.

Schmid has suggested what amounts to a facies model for cave sediments:

In a cave which does not lie in the flooding area of a stream, lake

or sea and cannot be flooded by water from inside the hill, we can expect the various parts of the cave to have accumulated the following materials:

Caves possessing a sill will collect there scree material from the brow of the cave. The scree can contain stones, soil and plants and animal remains from the entire area of the hill above the cave. In times of strong winds, during loess formation for example, there will be present a considerable amount of aeolian deposit. The sediment, in other words scree plus material blown in, is then subjected to immediate processes of soil formation, to which plants contribute. Animal life appears, most clearly evidenced by snails. The swifter the process of deposition, the less the intensity of soil formation. Marked accumulation of scree, which in itself has been exposed to several processes of decomposition, prejudices a direct climatic interpretation of the sequence of layers. Another difficulty is the likelihood of part of the sediments having been washed away.

The area under the brow of the cave is called the entrance. At the entrance is the drip-line of rain or melted water running off the brow of the cave. Scree from above falls on this entrance area, as well as on the sill, in varying amounts. Material is also blown in this far. To these items are added decomposition products from the front of the roof of the cave. Since this area is under the unimpeded influence of climate, we can, by means of the deposited material, directly deduce the climatic causes of decomposition. During undisturbed sedimentation, material is in the main accumulated in the entrance area, because here scree plus blown-in material and decomposition products unite, and above all the influence of the weather is most active. In the event of sedimentation being not too rapid, plants and animal life develops as well. Inclusion of soil particles and organic remains from the scree is also a possibility.

Sediments thus reveal the climatic influences which create decomposition, together with those influences which produce soil formation. This being a contemporary occurrence, we are provided with a basis for reading climatic conditions by means of sedimentary processes and thereby both deducing relative dates and reconstructing environments.

A drip-line is only formed when the brow of the cave is of such a shape that rain and melted water can run along it. The drip starts where the brow meets the roof of the cave. Most caves possess such a drip-line. The water which drips or sometimes pours onto the floor of the cave loosens all fine particles and carries them away. This results in a strip of washed-off coarse material being exposed. If loosely packed stone material lies underneath then the water carries the fine material into the interstices. Thereby coarse and splintered rubble lower down receives 'secondary' earth. More frequently, however, the fine materials are deposited in front of and behind the drip-line. If the water dripping from the brow of the hill contains calcium carbonate, which it most commonly does, then this is deposited on the floor as sinter. On the other hand water containing CO_2 can also dissolve calcium carbonate, which it deposits immediately adjacent to or underneath as sinter; the more hollow spaces the percolating water finds, the deeper the calcification of the sediments. Both processes generally occur together. In a cave whose brow is eroded back we can deduce the one-time position of the entrance from a similar calcification of sediments in front of the present brow of the cave.

In short, whenever the structure of a sediment makes washing a possibility we must examine minutely whether earth and travertine are primarily or secondarily deposited. On the sill, washing of this kind is caused by rainfall; at the entrance and further inside by dripping water.

> In the front part of the cave sedimentation is directly influenced by the weather outside. But little light enters. There is a lack of vegetation, snails and soil formation. The rubble from the roof and walls settles down on the cave floor and is subjected to the same climatic influences as at the entrance, though the effect is less. There is, however, no extraneous material. For that reason the climatic conditions of sedimentation can generally be easily read in the front part of the cave. In the interior of the cave sedimentation is subject to different laws. Daily and seasonal changes in climate do not penetrate or at least not directly. Therefore only important long-term changes can alter the mode of sedimentation. Earth and rubble can be included here, as evidence of marked changes in the course of major climatic events.
> If we find materials in the lower sediments of the cave entrance which could only have been created in the front part, or even the interior of the cave, then they must have been secondarily carried to the entrance, in which case we must examine whether water activity or solifluction is responsible. Alternatively, if their creation can be proved as autochthonous, then we have proof that the mouth of the cave has been shifted back by weathering of the hillside. Thus we are now looking at what was once an inner section of a longer cave. (1969a:155-6)

This model is extremely crude and the present author would disagree with a number of the suggestions. Much too much weight is given to climate and the whole model is too static and compartmentalised. Consider, for instance, the effect which the slope of the sediment surface might have on this model, and then consider why a surface should be sloping at a given angle and in a given direction in the first place. Nevertheless, the model represents a real attempt to see the cave in three dimensions, with sediments representing processes and environments in time and space. The work of Schmid (cf. 1958, 1969b) has always been exemplary in this respect.

Little progress has been made beyond Schmid's model, although work on specific cave sites has sometimes shown increasing insight. Other general models include that of Sutcliffe (1970), which covers the broad relationships between bone material and cave deposits, and that of Ford (1976, fig.2.13 and relevant text), which covers the sediments to be expected in a total karstic system. In the present text, several cases of motifs and cycles are mentioned, with discussions of the processes and environments

which may, theoretically, have given rise to them. Facies models, as such, are not proposed because it is felt that the modern comparative data are not yet sufficient.

3.3. A Case Study of Sediment Patterning: Three Cantabrian Caves.

Butzer (1981) discusses in some detail the sequences from three Spanish caves, Cueva Morin, El Pendo and El Castillo. The paper illustrates well some of the aims, possibilities and difficulties of cave sedimentology. It also gives us the best available opportunity to observe a case of patterning although, unfortunately, not at the structural level.

The study of these three caves is well supported by a consideration of the region as a whole, including data on geomorphology, sediment bodies (with some structural information) and other cave sequences in the area. The sedimentary environment in each of the three caves, at each level, is discussed, with due recognition of the significance of major structural properties (eg. cryoturbation structures) and of more or less autochthonous sediments, such as travertine and limestone scree. The really novel part of this paper deals with what Butzer terms "colluvial influx":

> The preceding evaluation of external and internal sedimentation processes [at Cueva Morin] implies that decalcified residues, as processed here, will overwhelmingly record external sediment introduced via the cave entrance. For one, the cave has been little enlarged during late Pleistocene time, judging by the absence of rockfalls in the excavation area, while the amount of detrital sediment released through decalcification of almost-pure limestone within the cave must have been minimal. Secondly, the roof is not fissured, and in the past water has dripped in at a very few points, below intersecting master-joints. In addition, Holocene sediment has accumulated exclusively at the cave entrance, while Pleistocene deposits of the interior are locally and minimally capped by a millimetre or two of calcite, but no detrital sediment. This precludes a measurable component of clay or silt entering

> the cave by roof cracks, rather than from the entrance. Consequently, the decalcified cave sediments should provide a temporal record of changing ground cover and slope runoff energy, filtered by the mechanics of moving sediment diagonally across the irregular rock outcrops on either side of the cave mouth, or directly down the cliff face above the entrance. Two interrelated processes are therefore involved: (a) erosion and transport on the external slopes, and (b) runoff energy between the drip-line and the cave interior. (Butzer 1981:149-50)

Butzer is therefore proposing to examine a textural property (aspects of particle size) of a component of the sediment of each defined bed (the decalcified residue), in order to deduce something of the mechanics of sediment mobility in the surrounding landscape. The aim and reasoning appear totally valid.

> It should be possible to estimate these processes from the textural data, using the prototypes provided by sediments from the Morin quarry exposures. Mean grain size (Mz) should reflect external erosion and transport in a generalized way, providing an indirect parameter for slope energy. On the other hand, sand content should record peak runoff energy along the internal axis of the cave, while clay levels would probably reflect median or waning energy conditions. Finally, sorting (So) should be sensitive to the periodicity of hydrodynamic energy as well as to the potential input of mass movements. (ibid., p.150; emphasis added)

Butzer is using well established general mechanical theory to isolate parameters that must be measured to achieve the stated aim. However, he is also proposing to use "prototype" sediments from the vicinity of Cueva Morin. Ideally, Butzer would need a series of prototypes which are demonstrably reacting to different energy levels in the modern environment. But, as so often happens, such prototypes are not readily available. Therefore, Butzer uses what material is to be found on the local hillslopes, on the agricultural flats and in small fissures and caves exposed in a nearby quarry. Since none of these deposits are unequivocally linked with modern energy levels, they themselves must first be interpreted. Totally circular argument is therefore a real danger: residue X in a quarry cave is interpreted as due to process Y and, because residue Z in the Cueva Morin looks like

residue X, residue Z is due to process Y. Butzer does not really fall into this trap, although the significance of this comparative material is somewhat overestimated, to the present author's mind. The usefulness of these sediments lies not so much in their validity as perfect 'prototypes', but rather in the way that they colour the theoretical mechanical model. The occurrence of various similar residues in the landscape suggests that no great problems of radically changing sediment source, giving misleading inherited properties, is to be expected in the caves. The range in values of the various parameters, and the range in associations of values of different parameters in the same sediment, lend a local 'flavour' to the study which serves as a rough scale against which the Cueva Morin sediments can be measured. Finally, the topographic occurrence of the various external sediment bodies does in fact provide a little support for the theoretical model. It is clear that Butzer understands these points, and it is therefore a pity that he chose the term "prototype".

> In order to investigate the nature of sample clustering, Mz was plotted versus percentage clay on one graph, So versus percentage sand and grit on a second. Those samples conforming to the textural definition of a loess were first singled out as a separate population. The remaining samples were then divided on both graphs into six sequential classes, representing a gradation of energy conditions. The 30 samples from El Pendo and 38 from Castillo were then added to the 72 from Morin, and minor adjustments of class clusters made. (ibid., pp.150-1)

The classification was subjective, based upon both theoretical considerations, 'scaled' by the local context, and intrinsic clustering behaviour of the samples. Butzer gives reasonable grounds for the inclusion of the El Pendo and El Castillo samples elsewhere in the paper. So that subjectivity did not produce unacceptable bias, the classes were tested by discriminant analysis. The results are fully reported (but note the displacement of values due to typographical error in Table 4)

and it is concluded:

> The programme identified 85.0% of the 140 samples as correctly classified, although all but nine are unambiguously defined on both of the original graphs. The classes are consistent and real. (ibid., p.151)

Seven main classes of residues were defined and interpreted; they will be briefly described here:

> Class 1 - Coarse, sandy and poorly sorted group. Periods of unusually high, overall hydrodynamic energy, during periods of exceptionally intensive rains, and at a time when surface soils had been stripped of all but their basal profiles.
> Class 2 - Coarse-grained polymodal, with repeated lenticular bedding [note this interesting structural property]. Fluctuating runoff velocities, despite high overall energy and intensive rains.
> Class 3 - Improved and consistent sorting. Sustained, moderately vigorous slope runoff and repeated, small-scale bursts of cave runoff. High silt content suggests loess as well as soil derivatives eroded during periods of heavy rains.
> Class 4 - Loess and loessic matrices (and silty flowstone residuals). Relatively vigorous runoff over loessic surficial sediment; penecontemporaneous with initial exterior loess deposition.
> Class 5 - Mainly suspended load, with soil derivatives and some loessic material; polymodal. Still quite effective external runoff.
> Class 6 - Fine material with high clay content. Mainly low-energy mobilisation within the cave, despite a measure of effective external runoff.
> Class 7 - Very badly sorted, very clayey sediments. Minor mass movement of the type observed on these Spanish hillsides today.
> (summary of Butzer 1981:151-3)

Butzer then states:

> As defined and interpreted here, the seven classes of residual facies are more than ordinal numbers. They also provide a numerical index of 1 to 7 that is inversely proportional to the vigour of external denudation and internal sediment mobilization. Runoff energy is a function of (i) the effectiveness of ground cover (rainsplash interception, rooting network, infiltration capacity), (ii) the available sediment (B-horizons, lithosols, loess) and (iii) the intensity of rainfall (dependent upon rate, duration, drop size, and wind speeds). Reasonable, even if highly general, inferences can be made for each of these variables, as argued above. The sedimentary facies class [!], in capitalized form (eg. Facies 1), is consequently used below as a quantitative index of both runoff energy (1 high, 7 low) and slope stability (1 low, 6 high; 7 with qualifications). As such it provides a semicontinuous measure of environmental processes and conditions that substantially transcend the cave micro-environment. (ibid., p.153)

The switch from "residual facies" to "sedimentary facies" is a dangerous misuse of the facies concept. Let us remember that we are dealing with a component of each sediment body, described only in terms of one group of textural properties.

Butzer goes on to use the residual facies and a significant number of other parameters (including C14 dates and archaeological assemblages) to set up a composite stratigraphy, combining the Cueva Morin, El Pendo and El Castillo sequences. This composite stratigraphy is couched in terms of climatic phases, which, in turn, are compared with the major phases recognised elsewhere in Europe. It is not the purpose of the present section to undertake an examination of the grounds for climatic reconstruction. What is of interest here is the possibility of patterning in the sequences of residual facies. Butzer himself makes the following comment:

> Apart from the important problem of temporal scale, a comparison of the three different cave contributions to the 44 phases of Table 7 [the composite sequence of climatic phases] shows that episodes with 'stable' external environment have only 57% the probability of representation that 'unstable' slope contexts have (21 stable phases recorded in 32 discrete horizons, 23 unstable phases in 61 horizons); this undoubtedly reflects slower sedimentation rates during times of reduced runoff energy. In other words, resolution improves significantly as the number of cave microstratigraphic columns is increased. Conversely, it is improbable that any one cave column will be reasonably complete. (ibid., pp.175-6)

First, we must establish a data base. Unfortunately, Butzer does not give tabulated primary data, so that the residual facies sequences must be extracted from his diagrammatic figures 10, 12 and 16. Because the actual position and number of samples are not shown, a unit of residual facies is here recognised either when a residual facies change is indicated or when there is a change of defined sediment body ("layer") without a residual facies change. This results in sequences of 44 (Cueva Morin), 29 (El Pendo) and 34 (El Castillo) residual facies units, 107 in all (fig. 11 in the present text). In the following, a 'higher' residual facies class will refer to a class nearer the end of the scale marked 1 (i.e. higher slope energy).

Butzer's statement concerning "phases", "horizons" and

Fig. 11 Cantabrian Caves - Residual Facies Analysis

Layer numbers for each of the three sites are indicated in brackets before the residual facies. See text for details. Data from Butzer (1981).

	PHASE	MORIN		PENDO		CASTILLO	
[youngest]							
	44	(29)	4	-		-	
*	43	(28)	5	-		-	
*	42	(27)	5	-		-	
	41	(26)	3	(22)	5	(18)	1
		(26)	2				
*	40	-		(21)	6	-	
*				(21)	7		
	39	(25)	3	(20)	3	(17)	3
		(25)	5			(16)	3
	38	(24b)	4	-		-	
	37	(24a)	5	(19)	5	(15)	6
						(14)	6
	36	(23)	6	(18)	5	-	
		(23)	4	(17b)	5		
		(23)	6				
	35	(22)	5	(17a)	3	(13)	1
		(22)	3			(12b)	1
*							
*	34	-		(16)	6	-	
	33	(21b)	3	(15)	4	(12a)	2
		(21a)	3				
	32	(20)	2	-		(11a)	3
		(20)	3				
	31	(19)	5	(14)	4	-	
				(14)	5		
	30	(18)	5	(13)	3	-	
*	29	(17b2)	5	-		-	
	28	(17b1)	4	(12b)	4	(10)	3
		(17a)	3	(12a)	4		
	27	(16)	3	-		(9b)	3
*							
*	26	(15)	5	-		-	
	25	-		(11b)	5	(9a)	6
	24	(14d)	3	(11a)	3	-	
				(10)	3		
	23	(14c)	3	(9)	3	(8b)	5

[continued]

Fig. 11 [continued]

PHASE	MORIN		PENDO		CASTILLO	
22	(14b)	4	-		-	
21	(14a)	3	(8)	3	(8a)	3
20	(13)	4	(7)	5	-	
19	-		(6b)	3	(7b)	3
18	-		(6a)	3	(7a)	3
17	(12)	2	(5)	3	(6b)	2
16	-		-		(6a)	3
* 15	-		-		(5e)	7
* 14	-		-		(5d)	6
* 13	-		-		(5c)	7
* 12	-		-		(5b)	6
11	-		-		(5a)	4
10	(11) (5)	1 1	(4b)	4	(4c2)	3
9	(10) (10)	3 2	-		(4c1)	3
8	(9)	1	-		(4b)	3
7	(8) (7)	2 3	-		-	
* 6	-		(4a)	5	-	
5	(6) (3)	2 2	(3)	5	(4a) (4a)	2 3
4	-		(2)	7	(3)	5
3	(2)	5	-		(2a)	6
* 2	-		-		(1)	5
* 1	(4) (1)	1 1	(1)	5	-	

[oldest]

"stable" and "unstable" slopes is based upon 'averaging' of the residual facies classes of the sediment bodies that have been correlated to define a phase. Basically what he is saying is that phases with higher facies have a better chance of being represented by sediment in more than one site than do phases with lower facies. This should translate roughly into a higher overall occurrence of higher classes: in our extracted data, 55 (51.4%) of the 107 residual facies units are of classes 1-3, whilst only 38 (35.5%) of them are of classes 5-7. Furthermore, it is logical to extend this argument to say that when part of the composite sequence is represented at one site only, and by lower residual facies, there may well be a hidden gap in the sequence. We will return to this proposition below.

Let us now examine the properties of the sequences at the three caves. If all possible gaps within the sequences are ignored, there are 104 recorded transitions from one residual facies unit to another, including transitions to the same facies across layer boundaries. These transitions are recorded in fig. 12. The first value in each box represents the number of times (n_{ij}) the row residual facies (i) is followed, in an upward transition, by the column residual facies (j). The second value represents the probability (p_{ij}) that, given the row residual facies (i), it will be followed by the column residual facies (j). The third value is a simple Markov parameter:

$$n_{ij}.\mathrm{In}(p_{ij}/p_j)$$

Where p_j = total of column j divided by the overall total of transitions (104).

The sum of the values of this parameter over all boxes gives the statistic:

	1	2	3	4	5	6	7	
1	0.429 3 +5.570	0.286 2 +2.183			0.143 1 -0.394	0.143 1 +0.399		7
2	0.200 2 +2.187	0.100 1 +0.041	0.500 5 +1.698	0.200 2 +0.786				10
3	0.054 2 -0.431	0.162 6 +3.140	0.460 17 +4.357	0.081 3 -1.533	0.216 8 +0.150		0.027 1 -0.368	37
4			0.385 5 +0.392	0.154 2 +0.263	0.231 3 +0.258	0.231 3 +2.634		13
5		0.044 1 -0.780	0.261 6 -1.863	0.261 6 +3.956	0.261 6 +1.248	0.130 3 +0.910	0.044 1 +0.121	23
6			0.300 3 -0.514	0.100 1 -0.300	0.300 3 +1.042	0.100 1 +0.041	0.200 2 +3.270	10
7			0.250 1 -0.354		0.250 1 +0.165	0.500 2 +3.301		4
	7 0.067	10 0.096	37 0.356	14 0.135	22 0.212	10 0.096	4 0.039	

Total transitions = 104

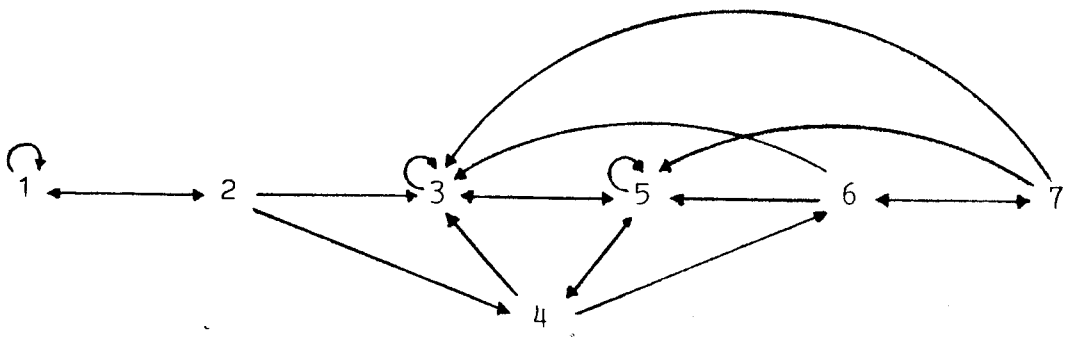


Fig. 12 Cantabrian Caves - Upward Transition Matrix

Stratigraphic gaps ignored; transitions without residual facies changes included. See text for details. Data from fig.11.

$$-2\ln\lambda = 2 \sum_{i=1}^{m} \sum_{j=1}^{m} n_{ij}.\ln(p_{ij}/p_j)$$

Where m = number of states (7).

The statistic $-2\ln\lambda$ follows the chi-square distribution with $(m-1)^2$ degrees of freedom. If the value of the statistic is significant, the sequences are behaving as Markov chains, that is, the transitions are not random. The flow diagram shows all the transitions that have a probability of at least 0.2.

In fig. 12 , the statistic takes a value of 65.15 with 36 degrees of freedom, which represents a >99% significant departure from randomness. Examination of the matrix, and of the flow diagram, shows that there is a strong tendency for a transition to be of only one or two classes (eg. facies 7 never jumps to facies 1, but is most likely to jump to facies 6). In fact, 40.4% of transitions are to a neighbouring class (a uniform distribution, with no Markov property, would have 26.5%), 23.1% are 2-class transitions (uniform 20.4%), 3.8% are 3-class transitions (uniform 16.3%), 2.9% are 4-class transitions (uniform 10.2%), 1.0% are 5-class transitions (uniform 8.2%) and none are 6-class transitions (uniform 4.1%). The matrix therefore shows a high degree of inertia, with mostly 'gentle' transitions. Indeed, the two greatest contributions to the Markov property (indicated by the highest positive values of the parameter $n_{ij}.\ln(p_{ij}/p_j)$) come from repetitions of facies 1 and of facies 3, and 28.8% of all transitions do not involve a residual facies change but merely a layer boundary (uniform 14.3%).

We may draw the following inferences from these figures. Even ignoring the random element that will be introduced by gaps in the sequences, there would appear to be a highly significant

	1	2	3	4	5	6	7	
1	0.750 3 +7.384	0.250 1 +1.022						4
2	0.167 1 +0.959	0.167 1 +0.618	0.500 3 +1.219	0.167 1 +0.266				6
3	0.035 1 -0.604	0.172 5 +3.238	0.448 13 +3.857	0.104 3 -0.623	0.207 6 -0.990		0.035 1 -0.108	29
4			0.300 3 -0.313	0.100 1 -0.247	0.300 3 +0.620	0.300 3 +3.207		10
5			0.294 5 -0.623	0.235 4 +2.430	0.353 6 +2.216	0.118 2 +0.272		17
6			0.125 1 -0.980	0.125 1 -0.024	0.375 3 +1.289	0.125 1 +0.194	0.250 2 +3.716	8
7			0.250 1 -0.287		0.250 1 +0.024	0.500 2 +3.160		4
	5 0.064	7 0.090	26 0.333	10 0.128	19 0.244	8 0.103	3 0.039	

Total transitions = 78

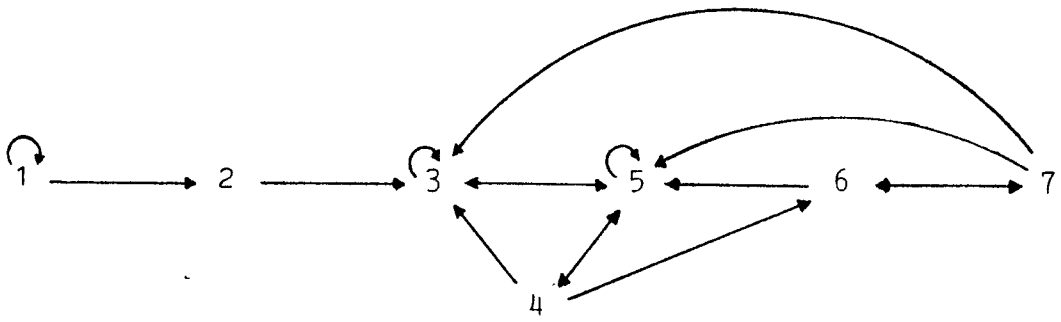


Fig. 13 Cantabrian Caves - Upward Transition Matrix

Stratigraphic gaps recognised; transitions without residual facies changes included. See text for details. Data from fig. 11.

	1	2	3	4	5	6	7	
1		1.000 1 +2.180						1
2	0.200 1 +1.661		0.600 3 +2.687	0.200 1 +0.163				5
3	0.063 1 +0.506	0.313 5 +5.094		0.188 3 +0.302	0.375 6 +2.554		0.063 1 +0.100	16
4			0.333 3 +0.921		0.333 3 +0.921	0.333 3 +2.776		9
5			0.455 5 +3.095	0.364 4 +3.045		0.182 2 +0.642		11
6			0.143 1 -0.538	0.143 1 -0.173	0.429 3 +1.681		0.286 2 +3.226	7
7			0.250 1 +0.020		0.250 1 +0.020	0.500 2 +2.664		4
	2 0.038	6 0.113	13 0.245	9 0.170	13 0.245	7 0.132	3 0.057	

Total transitions = 53

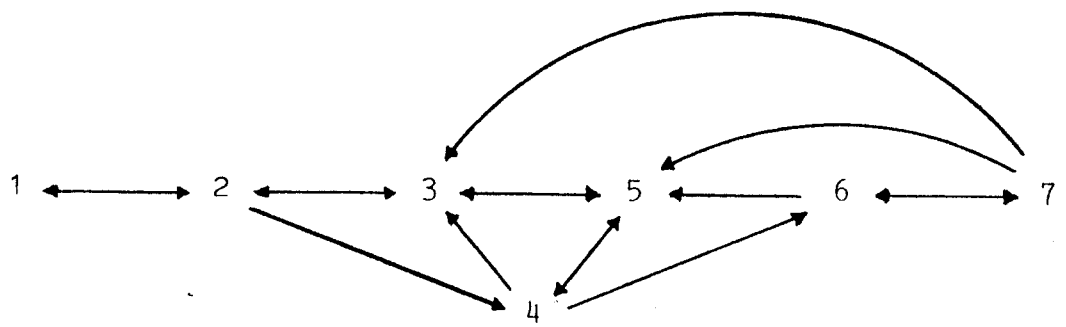


Fig. 14 Cantabrian Caves - Upward Transition Matrix

Stratigraphic gaps recognised; transitions without residual facies changes excluded. See text for details. Data from fig. 11.

number of 'gentle' transitions that are of the type one would expect in nature. That is, there are a large number of cases in which residual facies, and thus exterior slope conditions, changed only gradually. This would most probably occur if sedimentation within the caves had often been more or less continuous, with not too many periods unrepresented by sediment. Nevertheless, once a significant gap occurred, it may have been of considerable duration.

However, there are still some extreme transitions and Butzer's correlation of the three sequences suggests that gaps do indeed occur. If we accept the correlation, or at least the probable location of major gaps, we can construct two further matrices, one including transitions without residual facies changes (fig. 13) and the other showing only residual facies changes (fig. 14). As expected, the Markov statistic continues to give a >99% departure from randomness. Note how some extreme transitions (eg. 1-5, 1-6 or 5-2) have now disappeared; they were not real because they contained an unconformity. Of the remaining transitions (fig. 13), 32.0% involve no residual facies change, 41.0% are 1-class transitions, 23.1% are 2-class transitions, 1.3% are 3-class transitions, 2.6% are 4-class transitions and there are no 5- or 6-class transitions. Butzer's recognition of gaps therefore receives some support, although it is not impossible that he used extreme transitions to help define gaps in the first place.

The following table is a breakdown of how residual facies are placed with respect to identified gaps. Note that rows do not necessarily add up to the total because some residual facies units may have gaps both above and below them.

FACIES	AFTER GAP	CONFORMABLE	BEFORE GAP	TOTAL
1	4 (50.0%)	1 (12.5%)	3 (37.5%)	8
2	4 (40.0%)	3 (30.0%)	3 (30.0%)	10
3	8 (21.6%)	19 (51.4%)	11 (29.7%)	37
4	3 (21.4%)	7 (50.0%)	4 (28.6%)	14
5	7 (29.2%)	15 (62.5%)	5 (20.8%)	24
6	2 (20.0%)	7 (70.0%)	2 (20.0%)	10
7	0 (00.0%)	3 (75.0%)	1 (25.0%)	4

Clearly, the exterior slope energy is proportional to the percentage of occurrences after gaps, and inversely proportional to the percentage of apparently conformable occurrences within subsequences (i.e. without neighbouring gaps). It is tempting to infer that, when exterior slope energy was high, erosion was also occurring within the caves. However, we must not make the mistake of letting residual facies become 'sedimentary facies' (supra). Only structural information from the cave sequences could provide a reliable indication of whether or not this interpretation is likely to be valid. The percentages of different residual facies falling before a gap are reasonably constant, although higher classes show marginally higher percentages, probably because such residual facies units are often themselves followed, after the gap, by another high residual facies unit (cf. figs. 11 and 13).

We may now make the following suggestions concerning gaps in the sequences at Cueva Morin, El Pendo and El Castillo. Gaps may occur anywhere in the sequences, but there are three types of position that appear most likely:

(1) Where the composite sequence is based upon one cave only and the residual facies units are of low class;

(2) Where there is an extreme facies transition;

(3) Where there is no overlap between subsequences in different caves.

In fig. 11 , asterisks are used to show those points in the composite sequence considered to be most at risk. The construction of composite and regional stratigraphies is discussed in detail in Chapter 17. However, it will be interesting to note here that the points marked with asterisks usually coincide with some important 'event' recognised by Butzer. For example, Phase 40 is the Alleröd. The Gravettian takes over from the Aurignacian at the 34/35 boundary. Phase 29 marks the Würm II/III and the Mousterian/Upper Palaeolithic boundaries. Phase 26 is a Würm II mild phase. Quina Charentian first appears in Phase 16, which is also the start of Oxygen Isotope Stage 4. Phases 12 and 13 are in the middle of the Würm I/II interstadial. The 6/7 boundary marks the end of Oxygen Isotope Stage 5e. The 1/2 boundary is equivalent to the Oxygen Isotope Boundary 6a/5e. One wonders how much our view of the past is based upon discontinuities in the sediment record.

The above analysis represents only the 'weakest' form of information on patterning that could be extracted from such cave sequences as these. Much more interesting discussion would result from similar treatment of such data as true sedimentary facies sequences, the relationship between residual and sedimentary facies sequences, or the relationships between thickness of sediment bodies and various facies types. Although some data of this kind are presented by Butzer, the data collection stage becomes critical in analysis of complex patterning and it is rather dangerous to build inferences on information that was not collected with this specific aim in mind.

4. FIELDWORK

The preceding chapters have covered the historical setting for cave sediment research, and introduction to the cave environment and a discussion of a wide range of sediment properties and their significance. Before continuing with more specific topics, we must pause to consider the general relationship between the researcher and his study material.

The subject of fieldwork is mentioned over and over again in this thesis. This short section has been given the status of a separate chapter in order to underline the importance which the author attaches to fieldwork. Few researchers would disagree with this general proposition but, at least with respect to archaeological sedimentology, the most common attitude towards fieldwork is distinctly ambivalent.

Shackley states:

> A specialist [sedimentologist] cannot work isolated from the site. He needs to see the problems appearing and to advise and comment on recording and fieldwork. [...] A surprising amount of information can be recovered from careful field descriptions, which saves a great deal of wasted laboratory time and speeds up the appearance of excavation reports. [...] In many cases standard record sheets can be designed before the excavation, to cope with routine recording, and specialists need only be called to the site to deal with particular problems. Since the question of sample-taking is more delicate than is often supposed, most specialists will prefer to take their own.
> [...]
> There is a marked division of opinion about the merits of sediment field recording systems, a point which was apparent at the 'Sediments in Archaeology' symposium held at Southampton University, December 15-16, 1973 (Davidson and Shackley, 1975 [published in 1976]). One school of thought believed that it was essential to record the selection of relevant sediment characteristics which was thought at the time of the excavation to give the maximum potential information yield, and the other school argued that this was useless, since the information recorded

> would always be obsolete by the time that the excavation finished, because the questions to be asked would have changed. The writer holds the former opinion.
> However, it is always difficult to decide what sediment characteristics to record, just as it is difficult to judge when a sample is needed. The recording criteria will vary with the nature of the site, the sets of problems posed and the competence of the operator, as well as with the anticipated use of the results. [...] Various manuals have been produced for standardising field notes, the most useful being that published by the Soil Survey of Great Britain (1960), soon to be superseded by a new addition (Avery, 1974 [entitled Soil Survey Laboratory Methods]). This contains a methodology for recording a soil profile, which can usually be adapted for the qualitative recording of an archaeological sediment sequence.
> General information to be noted includes the precise position, site name, grid reference and locality, which can be printed onto a standard form. The date, feature or stratum number and the identity of the operator are always required, together with the shape and dimensions of the layer under consideration and its main characteristics. The nature and mode of formation should be recorded, and a description of the sediment texture. Several ... chemical tests ... may also be used in the field. (1975:10-11)

Turning to the proceedings of the symposium mentioned, there is no sign of the discussion of the relevance of field recording. However, there was some discussion of sampling.

> Mr. G. de G. Sieveking suggested that sampling was not always useful and quoted the excavations by Dr. Pales at Malarnaud-Soulabe where all the sediment from the cave had been kept. The speaker [F. G. Fedele] commented that sampling was not an exclusive means to record sediments and that it must be formally derived and supplemented by field observation. In certain cases total removal could be justified. Professor A. C. Renfrew then suggested that it might be simpler not to excavate the site! (Davidson & Shackley 1976:48)

> Following a point made by Professor A. C. Renfrew, the speakers [J. A. Catt and A. H. Weir] stated that in their opinion it was always necessary for a specialist to visit a site and take his own samples, and they emphasised the need for archaeologists to pose specific questions rather than generalisations. These views were supported by Professor V. B. Proudfoot and Professor G. W. Dimbleby who were of the opinion that specialists working on a site should always collaborate in sampling so that conclusions could be integrated. (ibid., p.91)

Catt and Weir make the following point in their personal contribution to these proceedings:

> Archaeologists have realised the importance of studying sediments and reconstructing past sedimentary environments, but they have usually relied almost entirely upon field observations. These are of course still fundamentally important, but geologists have found that the fullest understanding of sediments depends also upon a range of laboratory techniques, many of which demand specialist knowledge or apparatus. (1976:65)

What is the gist of the above quotations? First, there

are two sorts of people: specialists and site operators. Specialists should advise, but operators may do the field recording. Some people think field recording is useless, because changing objectives transcend all primary data collection; others do not, apparently because it "saves a great deal of wasted laboratory time and speeds up the appearance of excavation reports". It is difficult to decide which criteria to record (specialist advice needed?), but one factor in the choice of criteria will be the competence of the operator (to choose or use criteria?). Pedological (laboratory?) manuals are the most useful guides to sediment description in the field. When an operator has put his name, the date, etc., on a standard record sheet, he should describe a numbered stratum (defined by whom?), not forgetting to jot down a note or two concerning the "mode of formation" of that stratum. Sampling is a "delicate" operation (unlike field observation?), best carried out by the specialist, but there must be "collaboration" so that conclusions can be "integrated", presumably because otherwise there would be no link between the descriptions provided by the site operator and the sampling positions chosen by the specialist. All in all, field observations, which are "still fundamentally important" (why?), "supplement" sampling. Observation is all very well but "the fullest understanding of sediments depends also upon a range of laboratory techniques, many of which demand specialist knowledge" (and field observation does not?).

The present author is not suggesting for a moment that any of the named researchers should seriously accept the above analysis of their statements. After all, everything has been twisted to make it sound ludicrous. So what do these statements really mean? Read them any way you will; the ambivalence will

not go away.

The question of how to do fieldwork is irrelevant and cannot be salvaged by reference to any number of manuals. The only pertinent question is why do fieldwork. The simple answer is that sediments are to be found in the field, not in a laboratory. The laboratory is merely a comfortable place in which to do some of the more fiddly jobs that may require fragile or cumbersome equipment. Laboratories are useful, but they constitute a compromise, since many of the more interesting properties of sediments are not 'transportable' (cf. Chapter 3). Even the use of box samples, lacquer peels or plastified columns very soon runs into problems of scale. This simple fact is recognised by most geologists:

> The most fruitful approach is the use of the vertical profile and the geometry of sediment bodies. This approach to the problem of environmental analysis implies a return to the field and less dependence on laboratory analysis of samples. It requires careful measurement and study of stratigraphic sections and the mapping of sedimentary bodies. (Pettijohn 1975:542-3)

There is some confusion in the literature concerning the concepts of 'fieldwork', 'observation', 'sampling', 'laboratory work' and 'analysis'. Fieldwork and laboratory work are carried out in different places, as the terms suggest; there is no implication that one type is intrinsically better, or more accurate, than the other under all conceivable circumstances. The terms 'observation' and 'analysis' are often used as synonyms for 'qualitative method' and 'quantitative method'. Alternatively, the two terms may be considered as the end members of the continua 'simple-complex' or 'more or less quantitative'. Usage has not been strictly defined and there would be no advantage in doing so. Analyses are not restricted to the laboratory, nor are observations restricted to the field. There

is often an implication that analytical (quantitative) data are 'better' than observational (qualitative) data, a proposition that will be examined in Chapter 6. A sample is a part considered to be representative of the whole, with respect to the property of interest (cf. section 6.2.). Observation calls for sampling just as often as analysis; a sample may never leave the site and subsampling is often necessary in the laboratory. The choice of a site is an exercise in sampling, as is the positioning of a section. Whether the operator is wearing a hat marked 'specialist sedimentologist' or 'archaeologist' is of little importance. It is the competence of the operator with respect to the task in hand that counts. The execution of laboratory analyses requires special skills and knowledge, but these may be acquired by anyone with very little effort. The expertise required for sampling, field description, choice of observations and analyses, and interpretation of the sediments - factors which represent a unified body of reasoning which cannot be compartmentalised - is only acquired after considerable experience. Such experience is gained mostly in the field by cumulative observation of ancient sediments and modern sedimentary environments. It is a contradiction in terms to state that a particular operator is highly experienced in describing sediments (or even in drawing sections) but has little experience in interpreting them. Similarly, if an inexperienced operator has carried out the field description, no amount of expertise on the part of the 'specialist' in the laboratory will suffice to strengthen that weak link.

The ambivalent attitude towards fieldwork has nothing whatsoever to do with friction between site operators and 'specialists', although academic demarcation disputes may often cloud the issue. It is up to the competent operator to demonstrate

his reasoning. Laville states:

> Les études effectuées sur le terrain sont la base de notre travail, les analyses de laboratoire n'ayant en fin de compte d'autre objet que de nuancer, préciser ou démontrer les hypothèses et les interprétations tirées de la seule observation des dépôts. (1975:29)
>
> [Trans.: Field studies are the basis of our work, laboratory analyses having essentially no other object but to qualify, detail or demonstrate the hypotheses and interpretations derived solely from observation of the deposits.]

This is the ultimate statement on this subject. For his part, the present author would like this to be true of his own work, but it must be admitted that practice has often fallen far short of this ideal. Laville makes this claim for his work, yet he fails completely to demonstrate its accuracy in the body of his text. We are left with the feeling that perhaps, once again, only lip-service is involved.

> Whether thorough field observations are undertaken or not, it is becoming more and more rare for such observations to be published in detail, probably because of the elaboration of laboratory techniques and the growing mystique of mathematical manipulation. (Collcutt 1979a:291)

The scientific method has often been conceived in the past as a distancing of the observer from the thing observed, a sublime state of objectivity. Modern scientific thought has recognised the dangerous fallacy in this presumption. To observe is to change, since the only way to learn is to watch how the subject reacts when we do something to it. The more removed (spatially or conceptually) a sediment becomes from its field occurrence, the more it is changed. If the reader objects to this philosophising by stating that such properties as mineral composition are not likely to be changed merely by the act of removing sediment to the laboratory, then let him explain why we find it necessary to employ sampling theory if not to try to compensate for just such a change. The minerals may not have changed but the subject, the mineral composition of which we

wished to observe in the first place, most certainly has. Of course, fieldwork, even at a minimum level of intervention, also involves changing the sediments. The author is not suggesting that all study is thus doomed to failure. It is simply that the observer must become more self-conscious in order to evaluate the effect he is having on the sediments at each stage of study. All reasoning must take as its starting point the most undistorted evidence available to us - the field exposure. This first encounter with a sediment is the point where the greatest experience is both required and acquired.

5. BOUNDARY FEATURES AND STRATIGRAPHY

5.1. Space and Time

In section 3.2., the equivalence of lateral variation at a point in time and vertical variation through time was discussed in the context of facies models. Sedimentary stratigraphy has its own relativistic principles and space and time will often be found to be interchangeable. Thus, the phrase 'sedimentary sequence' implies both a physical stack of sedimentary bodies and temporal ordering of these bodies.

Collinson and Thompson discuss the basic principles that are used in sedimentary stratigraphy.

> The principle of original horizontality of beds: that most beds were once laid down either parallel to the earth's horizontal or at very low angles. Exceptional beds deposited at angles up to 40° may be recognisable by their sedimentary structures. (1982:9)

This principle derives from the fact that the great majority of extant sediments and sedimentary rocks were laid down under water. For our purposes, it will be useful to produce a more general statement. The angle at which a bed was originally laid down was a function of the depositional process (medium and energy) and the properties (mostly textural) of the sediment. Application of actualistic principles will show whether the observed angle diverges from the probable original bedding angle.

> The principle of original continuity of beds; that groups of beds are normally laterally extensive and maintain their thickness and continuity for great distances. Individual beds and layers are just as likely to be lenticular over short distances. (ibid.)

Again, we may generalise. The original extent of a bed is governed by the original extent of the depositional environment, often defined by a change to a non-depositional or erosional environment. Application of actualistic principles will show whether the observed extent of a bed is significantly less than its probable original extent.

> The principle of superposition of beds: that the younger beds are deposited on top of the older beds in a sequence. (ibid.)

Superposition is not necessarily perfectly vertical or coterminous. Younger beds may be progressively shifted away from older beds (off-lap) or may progressively encroach over older beds (on-lap); these patterns are equivalent to prograding and retrograding sequences, respectively, in subaqueous contexts. More irregular patterns of lateral shifts may also occur. All that is necessary to demonstate superposition is to show that some part of a younger bed was laid down vertically above some part of an older bed. Care should be taken to identify the relevant scale of bedding. If two sediment units have a sloping contact, situations exist in which the uppermost subdivisions of the 'lower' unit are younger than the lowermost subdivisions of the 'upper' unit. The contact is said to be time-transgressive. For example, consider a pool, with deposition of fine sediments, bordered by a stalagmitic rim. For some reason, the pool shrinks through time and the stalagmitic rim encroaches over older pool sediments. If pool sediments and stalagmite were merely thought to be two separate beds, one might arrive at the false conclusion that all the stalagmite was younger than all the pool sediments. The correct relationships are only seen by examining minor sedimentary units within the two main units.

> The principle that the way-up of beds can be recognised by sedimentary structures that have characteristic attitudes with respect to the top

or bottom of a bed. (ibid.)

This principle is useful when there has been massive post-depositional deformation of beds. The reasoning is similar to that employed in the principle of original bedding angle ('horizontality'). In caves, this principle is most often applied to decide the way-up of speleothems and other consolidated sediments which may or may not have been disturbed.

The principle of included fragments: that fragments of an older bed can be included in a younger bed, but not vice versa. (ibid.)

This principle is a useful generalisation but is not strictly true. More exactly, we may say that fragments of a younger bed will not be included in an older bed as long as they are larger than any original or post-depositional voids in the older bed. This rider is important in the cave context, where openwork scree is very common.

The principle of cross-cutting relationships: that a feature which cuts across a bed must be younger than it. (ibid.)

This principle is similar to the principle of superposition and the same caveat concerning time-transgressive contacts is necessary.

There is one further principle, that is of considerable importance in caves, which we may call the principle of hanging beds: when a bed of consolidated (cemented, indurated) sediment is firmly anchored to the rock, it will be older than all unconsolidated beds that show a normal superpositional relationship with it, but it may also be older than all beds (consolidated or not) lying below it or younger than any other anchored and consolidated bed lying above it.

The above principles embody the concept of stratigraphic sequence (spatial and temporal). The study of the stratigraphy of a site is the first priority but there is an immediate clash

caused by the need for iterative reasoning. Reconsider the following statements already made in section 3.1. The identification of bedding is subjective, since there is an underlying assumption that bedding is due to primary sedimentary processes (p. 94). The nature, degree and rate of change in properties necessary to define a change in bed depend primarily upon the processes thought to be responsible (p. 95). These statements are true. Furthermore, we have seen that some stratigraphic principles call for actualistic inference. Now consider this statement from the International Stratigraphic Guide:

> Lithostratigraphic units [including beds] are recognized and defined by observable physical features and not by inferred geologic history or mode of genesis. (Hedberg 1976:31)

We are therefore faced with a logical paradox. In order to produce a descriptive framework that is as free of inference as possible, we must actively refrain from considering sedimentary process. Such a framework is indispensable, both as a starting point for interpretation of the sediments and as an unbiased statement to be communicated to others. However, the recognition of a bed is inextricably linked with ideas of process. One solution would be to define two quite separate types of unit: lithostratigraphic beds and lithogenetic beds. But how would we then decide upon the degree of change necessary to define a lithostratigraphic bed? A proliferation of classification schemes is not the answer to this problem. The definition of lithostratigraphic units must be based upon changes in the sediments that our cumulative group experience tells us are likely to be significant under the widest possible range of sedimentary processes. We then compare each resulting unit with the set of processes we think might possibly have given rise to

that unit, in order to make sure that it will not be too general or 'coarse' for later lithogenetic interpretation. If it is too general, the unit may be subdivided as long as there are observable lithologic criteria (which there should be if we suspect a significant change of process within the original unit). If in doubt, it is probably better to define an over-detailed lithostratigraphy; units, and data from samples of these units, can always be grouped at a later stage, but later subdivision may invalidate earlier samples and observations. This state of affairs is tacitly recognised in the International Stratigraphic Guide:

> Because the many lithologic variations of rock strata offer a wide choice in drawing boundaries of lithostratigraphic units, the selection of these boundaries may be properly influenced by such other factors as lateral traceability, physiographic expression, fossil content, lithogenesis, and well log character, as long as the requirement of a substantial degree of lithologic homogeneity [within a defined unit] is maintained. (ibid., pp.38-9)

The Guide stresses the need for unit homogeneity, whilst the study of sedimentary processes needs a classification which will adequately represent change in the sediments. These are obviously the two sides of the same coin; a balance must be found where the requirements of both lithostratigraphy and interpretation are met. The present author cannot see the point of proving, in vacuo, that a unit is lithologically homogeneous and different from other units, merely on the grounds that intra-unit variation is less than inter-unit variation. Shackley (1972) has done just this (although the validity of the statistics used - correlation coefficients of inclusive graphic statistics derived from particle size analysis - is extremely dubious) with sediments from Badger Hole (Mendip), but such tests do not show how and why such units vary. The units may still be too coarse or too fine for any useful interpretation. Lithostratigraphy is of

primary importance but it is a tool, not an end in itself.

It is clear that, whether we are currently interested in lithostratigraphy or in environmental interpretation, zones of comparatively rapid change within the sediments must serve to define units. There are two main categories of phenomena of interest here and it is imperative that they be kept separate. The first comprises observable physical phenomena (changes in sediment) and the second comprises totally inferential phenomena (time gaps). There is considerable disagreement, if not confusion, over the nomenclature used for these phenomena; the author has chosen those definitions which seem most useful in the present discussion.

Collinson and Thompson state:

> The feature of measured sections most commonly ignored is the nature of contacts between units. (1982:177)

We may define such a contact as a <u>boundary</u>. Viewed objectively, a boundary may often have three-dimensional form and structure in its own right, in which case it is a <u>gradational</u> boundary. The gradation is between the sediment of the underlying bed and that of the overlying bed, and may be continuous (eg. a sand bed underlying a silt bed with a continuous fining-upwards gradation within the boundary zone) or interstratified (eg. sand below silt, with a boundary zone composed of interstratified but discrete sand and silt laminae, perhaps with sand laminae thinning upwards and silt laminae thickening upwards). For lithostratigraphic purposes, it may be necessary to define an arbitrary (but still observable) level within the boundary zone as the limit between units. Alternatively, a thick boundary zone may be given unit status in its own right. As a boundary zone thins, it will eventually be more useful to

consider it as an abrupt boundary, or bounding surface. A boundary zone may vary laterally in thickness, form (planar, undulating, irregular, etc.) and angle and direction of general dip. An extremely irregular (contorted) boundary may even show detached pockets of one sediment floating in the other bed. Bedding plane features occurring at a boundary may properly be included in its description. Obviously post-depositional features of the underlying unit (eg. alteration, cementation, burrows, faulting, etc.) may be included in the definition of a boundary only at a descriptive level and only if (a) the phenomenon never crosses to the overlying unit, (b) the phenomenon is never observed to be separated from the boundary by unmodified sediment, and (c) there is not the slightest suspicion that the phenomenon could possibly have been developed laterally rather than from a free sediment surface. A certain amount of inference is called for here to judge whether a phenomenon really was developed from a surface that post-dated the deposition of the extant lower unit and pre-dated the deposition of the extant upper unit. Clear truncation of a post-depositional phenomenon by the boundary does not necessarily disqualify it from inclusion in the definition of the boundary. The use of all post-depositional phenomena in boundary definitions calls for caution and arguments based upon extensive exposures. It may also be useful to record the contrast (visibility) of a boundary, since the sediments above and below it may vary laterally, causing the boundary to appear more or less distinct. Indeed, a boundary may not always be confined between the same pair of beds. Many boundaries are most apparent because of changes in the fine matrix between beds. However, if the matrix remains more or less constant, with only the coarse elements changing

(in frequency, composition, form, orientation, etc.), the boundary is said to be welded.

As was said on p. 96 , boundaries (and bedding plane features) may be characteristic of particular sedimentary processes. Nevertheless, it is still necessary to develop a purely descriptive vocabulary for lithostratigraphic purposes. Occasions may arise in which inference suggests that a boundary is the result of a major process (erosional or non-depositional) which has had no part whatsoever in the formation of the beds above and below. Major post-depositional changes of underlying beds may also be stratigraphically associated with a boundary (supra). In the present author's opinion, it may then be valid to allocate the boundary special status (together with an unequivocal name reflecting its own physical properties) and to include it in the stratigraphic column. This is not general stratigraphic practice and therefore such a step must be accompanied by detailed justification.

The concept of unconformity has been linked with inference concerning possible time gaps between beds. Collinson and Thompson state: "An unconformity is a break in a stratigraphic sequence resulting from a change in conditions that caused deposition to cease for a considerable time."(1982:17). Thus, not only is a major time gap (time not represented by extant sediment) suggested, but also a period of non-deposition. This has been the consensus ever since the time of Hutton. Tomkeieff (1962) has discussed unconformity and has recognised the complexity underlying any definition, including, on the one hand, inference with respect to time and process and, on the other, pure observation of the structural relationships in the sediments themselves. The present author is extremely wary of

any classification that entails an a priori link between observation and interpretation. It is incomprehensible that such a classification should have survived the proper division between primary data and inference that has taken place in most other branches of geology.

Fig.15 shows schematically some of the many possibilities of structural relationships with and across a boundary. The relationships are defined by the following factors: (a) the form and dip of the boundary; (b) the angle and curvature at which the bedding planes of the upper and lower units meet the boundary; (c) whether or not the form of the bedding planes of the upper and lower units are demonstrably linked with the presence and form of the boundary. Any one of the relationships may be defined verbally, using simple geometric terms, without it ever being necessary to involve inferential arguments. For instance, it should not be difficult for the reader to identify the entry in fig.15 which corresponds to the description: upper unit with irregular, parallel bedding, mirroring an irregular but generally horizontal boundary, which truncates a lower unit with curved (concave-down), parallel bedding. This property of contiguous sediment units may be called contact geometry and is a wider concept than boundary geometry alone. Contact geometry may provide valuable clues concerning erosion events, periods of non-deposition, erosive processes, changes of direction of sediment input, time-transgressiveness of boundaries, deformation, and the like. Some of the entries in fig.15 correspond to the diagrams traditionally used to illustrate the structural components of the various types of unconformity. However, a given contact geometry is not proof of a given sequence of events and every entry in fig.15 has more than one possible explanation.

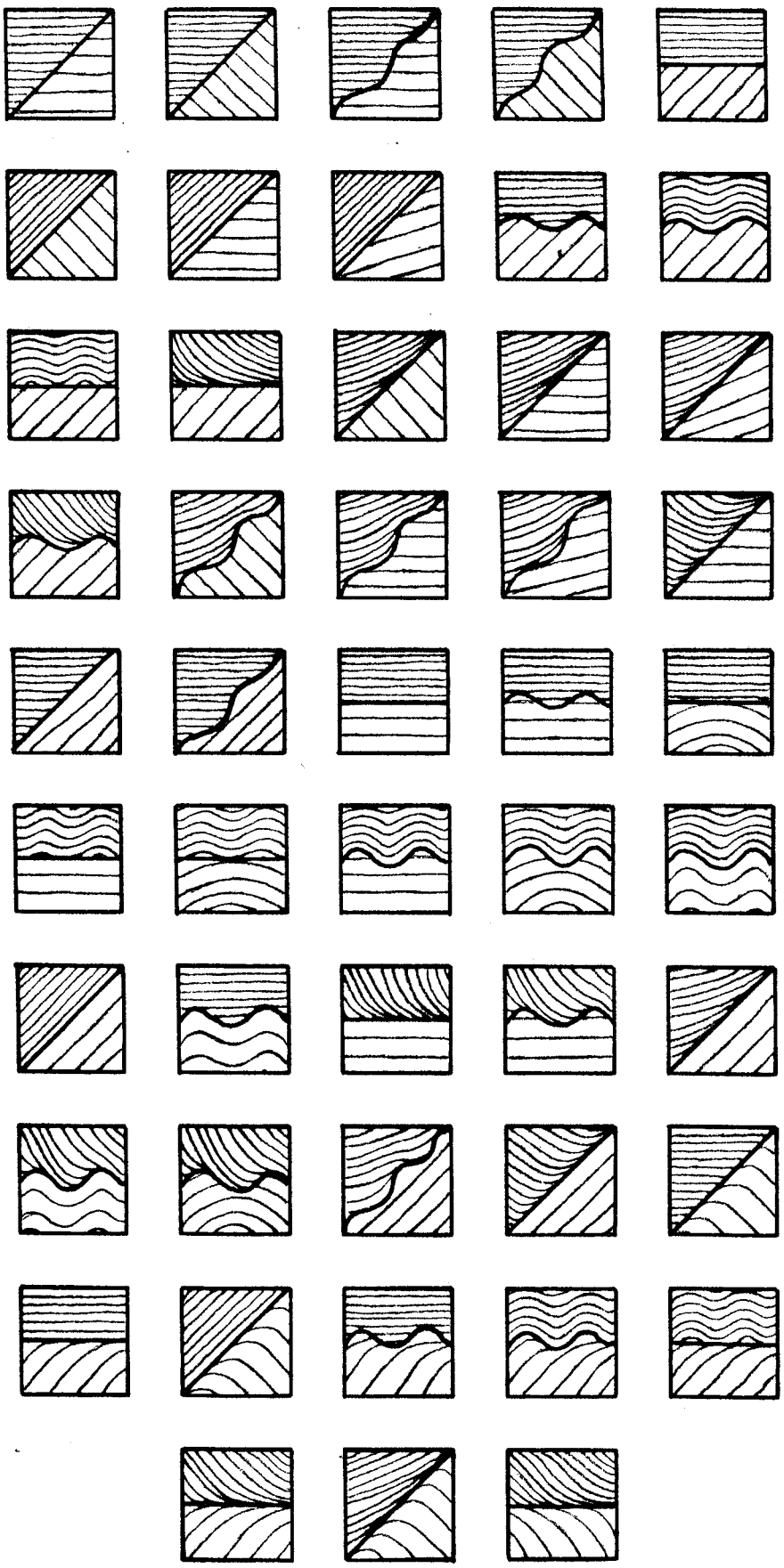


Fig.15 Contact Geometry

When a lower unit is truncated (i.e. when the principle of original continuity of beds suggests that a part of the unit is missing), we have very strong grounds for believing that an amount of time, unrepresented by sediment, separates the upper and lower units. We know that part of the lower unit is missing, and we may sometimes be able to deduce its original extent. However, what if other units, now totally missing, were also once present? Even when there has been no obvious truncation of the lower bed (as when the boundary and the bedding planes of both lower and upper units are all parallel), there is no guarantee that a time gap is not present. A gap may be due to erosion, non-deposition or a complex combination of the two, perhaps including unevidenced periods of sedimentation. There are only two generally accepted categories of time gap: diastems and unconformities. A diastem (also known as a 'non-sequence') is a break in deposition of short duration and/or local extent. An unconformity is a gap, traceable over wide areas, involving a major break in deposition (and often, in practice, a phase of erosion; supra). One may speak of 'major diastems' and 'minor unconformities' in order to cover a wide range of predominantly non-depositional gaps. There is no generally accepted term for a gap produced by erosion, probably because of the difference in time scales between quaternary sediment bodies and the earlier sequences studied by most sedimentologists. In pre-quaternary terms, erosion that removes significant amounts of sediment is almost always associated with, or follows, long periods of 'non-deposition' (i.e. comparatively slow terrestrial deposition and erosion) which are dependent upon initial emergence of marine sediments. The term 'hiatus' has often been used in the literature to indicate a gap without precise

connotations of duration or process, but the very imprecision of this term may lead to misunderstanding. Although any classification scheme based mainly upon ways in which physical evidence is absent will always be somewhat risky, it would appear that the subject of time gaps is long overdue for reconsideration.

5.2. The International Stratigraphic Guide

The International Stratigraphic Guide has already been mentioned above. The Guide is the product of twenty-four years of work by the International Subcommission on Stratigraphic Classification, a body "including almost every national and multinational stratigraphic organization in the world."(Hedberg 1976:2). We may conclude that the Guide deserves our attention. It is stated: "Many special or newly developing fields of stratigraphy, such as those dealing with ... the Quaternary, ... are barely touched in the present edition."(ibid., p.1). Nevertheless, the Stratigraphic Commission of INQUA has adopted the Guide as the basis for the description of quaternary lithostratotypes, and thus for lithostratigraphy in general.

It is very important that the nature of the Guide be clearly understood. A casual reader might mistake it for a mere code, to be applied only when one wishes to publish a formal definition of a stratigraphy thought to be of considerable national or international significance.

> The purposes of the Guide are to promote international agreement on principles of stratigraphic classification and to develop a common internationally acceptable stratigraphic terminology and rules of stratigraphic procedure - all in the interests of improved international communication, coordination, and understanding, and thus of improved effectiveness in stratigraphic work throughout the world. (ibid., p.1)

Such goals would appear to have little to do with the needs and objectives of a researcher involved in the study of a cave site, perhaps of only very local significance.

> The Subcommission offers its International Stratigraphic Guide as a recommended approach to stratigraphic classification, terminology and procedure - not as a 'code'. There is no intention that any individual, organization, or nation should feel constrained to follow it, or any part of it, unless convinced of its logic and value.
> [...]
> The Guide reflects the belief that the ideal concepts behind stratigraphic classification and terminology should be maintained, even if in practice the ideal can only be approached and not perfectly achieved. A relaxation of standards rarely leads to progress; and if compromises must be made in practice they should be recognized for what they are.
> Clarity of definition is as essential to stratigraphy as to any other science. The use of unreal concepts, vague and cloudy terms, and imprecise definitions is often defended because such procedure seems easier, simpler, more attractive, or more traditional than a more rigorous approach. However, if a concept or a term cannot stand attempts at precise definition, it is usually of dubious value. (ibid., pp.4-5)

The logic of lithostratigraphic classification will be the same, whether we are interested in the Grand Canyon or a minor cave sequence. Naturally, quaternary cave sites may present special problems which may require divergence from the recommendations of the Guide. We must therefore demonstrate the 'logic and value' of any such divergence. The declared purpose of the Guide is to promote international communication, so there is an understandable bias towards formal classification. However, if the procedures leading to a formal classification are logical and valuable, why not use them at an informal level as well? The present author recommends that, whenever possible, a site be approached as if it were to be the subject of a formal classification. When proper lithostratigraphic principles are applied from the outset, on site, the user is forced to question his own reasoning at every step and the chances of the unintentional inclusion of inferential arguments are greatly reduced. The Guide is not merely an arbitrary compromise for

the sake of communication - it is a powerful test. The author has found that, whenever he or a colleague has been tempted to ignore basic advice from the Guide in specific instances, sooner or later the 'better way of doing it' has been found to be concealing an initial error. When the error is corrected, there is no longer any difficulty in applying the procedures recommended by the Guide.

The remainder of this section is concerned with the application of stratigraphic principles to individual cave sites, with special attention to any problems that may arise from the declared intention of attempting to use the Guide at all stages and not just at the formal level.

The Guide proposes the following definition:

> Lithologic unit - a body of rock strata that is unified by consisting dominantly of a certain lithologic type or combination of lithologic types, or by possessing other impressive and unifying lithologic features. [...] The critical requirement of the unit is a substantial degree of overall lithologic homogeneity. (Diversity in detail may in itself constitute a form of overall lithologic unity.) Lithostratigraphic units are recognized and defined by observable physical features and not by inferred geologic history or mode of genesis [supra]. Fossils may be important in the recognition of a lithostratigraphic unit either as minor but distinctive physical constituents or because of their rock-forming character as in coquinas, diatomites, coal beds, and so on. The geographic extent of lithostratigraphic units is controlled entirely by the continuity and extent of their diagnostic lithologic features. Only major lithologic features readily recognizable in the field should serve as the basis for lithologic units. (ibid., p.31)

This definition presents no problems. Note how immediate the procedure will be, based as it is on observable data. Information gathered in the laboratory may of course be included, but only if the properties could be observed on site with relative ease. For instance, accurate measurement of alkali-soluble organic content in the laboratory may be used to provide comparative statements ('rich' or 'poor in organic matter') to be included in the lithostratigraphic description.

It is not the purpose of this section to discuss

biostratigraphy and chronostratigraphy in detail. However, it is necessary to examine the relationships between the three main types of stratigraphy.

> Rock strata are classified biostratigraphically by dividing them into units distinguished by differences in their fossil content. A biostratigraphic unit may be based simply on the presence of fossils as contrasted with their absence; on all kinds of fossils taken together or only on fossils of a certain kind; on the complete assemblage of fossils characterizing a certain stratigraphic interval or only on selected taxons; on a particular natural association of fossils; on the range of a fossil taxon or fossil taxons; on frequency and abundance of fossil specimens; on certain morphological features of fossils; on fossil indications of mode of life or habitat; on stages of evolutionary development; or on variations in any of the many other features related to the fossil content of strata.
> There are thus many different kinds of biostratigraphic units depending on the paleontological feature considered. Biostratigraphic units, like lithostratigraphic units, are relatively objective products of classification in that they are based on directly observable features in the rock strata. A biostratigraphic unit may be considered to be present only within the observed limits of occurrence of the paleontologic feature on which it is based. (ibid., pp.45-6)

This definition presents few problems with respect to eventual formal classification. However, there are certain general problems which must be considered with respect to the development of biostratigraphies, especially in the Quaternary. In practice, biostratigraphic classification is not immediate and depends upon extremely detailed and lengthy study which is not usually viable on site. Also, "while all sequences of layered rocks ... can be subdivided into lithostratigraphic units, biostratigraphic units can be recognized only in rocks containing fossil remains."(ibid., p.95). Since fossils usually occur as objects dispersed in sediment, they do not constitute a spatial continuum. There are therefore problems in the definition of what constitutes a 'rock stratum' in the biological context. In the Quaternary, there is very little evolutionary change in taxa, so that any assemblage must be interpreted to discover its significance. The application of taphonomic principles (cf. Chapter 14) is becoming increasingly accurate and capable

of distinguishing the complex blend of biological information represented in any stratum. However, taphonomy is already sufficiently difficult without adding the complications that would arise if the palaeontologist ignored lithologic boundaries. Lithostratigraphy is the only available starting point for the quaternary palaeontologist. It produces a sampling framework consisting of (lithologic) units which contain the least complex (the least ecologically and temporally mixed) biological assemblages which are readily accessible. This is not to say that the biostratigrapher is in any way constrained to remain within the lithostratigraphic framework. Several lithostratigraphic units may be grouped to form one biostratigraphic unit and there is no necessary correlation between the rank of the former (eg. beds or formations) and the nature of the latter. Furthermore, the lithostratigraphic framework is only a first approximation for the palaeontologist, and he should be encouraged to look for changes within a defined lithologic unit. Indeed, a biostratigraphic boundary may be properly placed within a lithologic unit or may even cut across a lithostratigraphic boundary (a comparatively rare occurrence in the terrestrial Quaternary). Lithostratigraphy is a stepping stone for biostratigraphy.

> Chronostratigraphic unit - a body of rock strata that is unified by being the rocks formed during a specific interval of geologic time. Such a unit represents all rocks formed during a certain time span of Earth History and only those rocks formed during that time span. Chronostratigraphic units are bounded by isochronous surfaces. The rank and relative magnitude of the units in the chronostratigraphic hierarchy are a function of the length of the time interval that their rocks subtend, rather than of their physical thickness. (ibid., p.67)

> Although lithostratigraphic and biostratigraphic units are largely established and distinguished on the basis of observable physical features, chronostratigraphic units are identified on the basis of their time of formation - a more interpretive character.
> Both lithostratigraphic units and biostratigraphic units are invaluable aids to developing a sound chronostratigraphic classification. (ibid., p.95)

It is clear that lithostratigraphy and biostratigraphy have precedence over chronostratigraphy. Because geochronometric methods (cf. Chapter 18) are so time consuming and expensive, it is extremely inefficient to apply them to all datable material. Furthermore, only a tiny proportion of the material in any stratigraphic sequence is even theoretically datable. Therefore, the stratigraphic relationships and significance of a sample must be ensured before it is dated. In practice, lithostratigraphy, biostratigraphy and many aspects of chronostratigraphy have precedence over geochronometry.

In summary, it may be said that lithostratigraphy is immediate, it is applicable to all the material of a stratigraphic sequence (biological and anthropogenic material included) and it is a useful (in practice, even necessary) framework for the development of other classes of stratigraphy. This implies procedural priority, not greater 'value'. Rather than fostering a sense of superiority, this realisation should serve to underline the responsibility which the lithostratigrapher must bear towards his colleagues in other disciplines (cf. p.568).

Reference was made above to the distinction between formal and informal classification and terminology. The Guide is rather too generalised for our purposes in its definition of 'informal':

> Informal stratigraphic terminology uses unit-terms only as ordinary nouns without the unit necessarily being named and without it being a part of a specific scheme of stratigraphic classification, for example, a chalky formation, an oyster zone. The initial letter of an informal unit-term is printed in lower case. (ibid., p.13)
>
> Informal lithostratigraphic units include lithologic bodies to which casual reference is made but for which there is insufficient need, insufficient information, or inappropriate basis to justify designation as a formal unit. (ibid., p.35)

The very object of informal systems is that they should

remain flexible. The following discussion must not be seen as an attempt to give formal meaning to informal terms. Rather, it is an attempt to distinguish the various levels of meaning which are important in any real application. Within the definition of 'informal' given by the Guide, we may pick out two main categories: (1) 'casual' terms, where there is no desire or need to use formal terminology; (2) 'obligatory' informal terms, where there is no possibility of using formal terminology, whatever our desires or needs. However, owing to the above declared intention of attempting to use the logic of the Guide at all stages, a third category must be recognised which must be both informal and also a part of a specific (explicit or implicit) scheme of stratigraphic classification. This category must comprise a system that is in as much as is possible homologous with formal classification, but which embodies the proviso that the system describes an interim state of knowledge. Merely as shorthand, such a system will be referred to below as preformal classification. The nomenclature employed may be of any type, as long as it cannot be mistaken for a formal system. For instance, the same terms, but in lower case, may be used (eg. bed > Bed, formation > Formation, etc.). If the distinction between preformal and casual usage is required, the preformal terms may be italicised (eg. bed > Bed, formation > Formation, etc.). Another method is to produce a parallel set of terms which may be italicised, with or without an initial capital (eg. layer > Bed, Unit > Member, Sequence > Formation, etc.). It should not be forgotten that the object of preformal classification is primarily the organisation of the reasoning of those researchers working on a given site, and not the communication of standard concepts to others. Nevertheless,

interim reports, and final reports where no formal classification is desired, may be published using preformal classification, whether or not the homology with the formal scheme is made explicit. The author feels that, in most cases, the homology should indeed be explained, simply because he believes that results can never be divorced from procedure. In those cases where the authors are seeking general support for raising the site stratigraphy to formal rank, preformal classification allows a period of discussion before the final decision. Since preformal classification, by definition, will often be subject to revision, synthetic texts should not refer to a unit or a site stratigraphy for which a preformal scheme has been explicitly used without a clear citation of the source publication or, at the very least, the author and year (cf. Linnaean taxonomy). In Part II of the present text, various kinds of preformal classification have been used, whenever a clear advantage could be seen to result; such usage is always explicitly recognised.

The various kinds and ranks of lithostratigraphic unit are defined on pp.32-36 of the Guide. The recommendations should be valid whether formal or preformal classification is required and a distinction will only be made below in specific cases.

At this stage, it is necessary to recall certain peculiarities of cave deposits. Such deposits are usually bounded on most sides (including above) by massive unconformities of extremely complex form, namely, the bedrock/sediment interface. The important point about this interface is that, whereas in many general geological situations reasonable assumptions can be made about the form (the 'behaviour') of an unconformity

beyond actual exposures, the bedrock/sediment interface in caves is often so irregular that extrapolation beyond visible exposures is very unreliable. Cave deposits are usually elongated bodies of sediment encased in hard rock, with large volumes of sediment being thus totally inaccessible. Even shelter deposits can only be approached from the outside. When dealing with linear sediment bodies, excavators, quite rightly, are disinclined to 'cut' the stratigraphy with rock-to--rock trenches, thus effectively restricting even further the available exposures. Two or more cave entrances in the same area may or may not be part of the same system, and may or may not have one continuous body of sediment. Energy fluctuations in caves, with high energy levels fragmenting pre-existing deposits and low energy levels often creating extremely localised deposits, usually result in rapid vertical and lateral variation in stratigraphic detail. Furthermore, isolated bodies of sediment, due to both natural causes and to the activities of past excavators, are extremely common. Many other types of site have similar problems but, in cave sites, these factors are particularly important. The problems will not disappear, even after extensive excavation and study. Some compromise will therefore be called for in the definition of lithostratigraphy in all cave sites. We may note the following generalities:

(1) Every effort should be made to map lithostratigraphic units of all ranks, but, in the cave context, it must be recognised that large volumes of any unit will usually remain inaccessible. A unit will be characterised by its lithology with the greatest possible care so that, if and when further exposures are uncovered, the unit may be extended with the

least possible disruption of the overall lithostratigraphic classification. Although standard excavation techniques provide the most satisfactory way of revealing stratigraphy, the possibility of core sampling should not be ignored. Jenkinson et al. (1983) have used a portable hydraulic ram system to core up to a depth of c.1.6m in Robin Hood's Cave (Derbyshire); they report that dolomitic limestone rubble was penetrated with little disturbance of the resulting core. Such a technique would prove an invaluable aid to extending known units, since a little, or even moderate, disturbance would not usually matter in this context. It should not be forgotten that useful lithostratigraphic units may sometimes be traced outwards, away from an entrance, and not just into the cave.

(2) Units of higher rank (eg. members or formations) should be initially defined in such a way that their lithology is stressed and less importance is attached to boundary stratotypes ("a specific point in a specific sequence of rock strata that serves as a standard for definition and recognition of a stratigraphic boundary" (Hedberg 1976:24)). In this way, when a new unit of low rank is discovered lying between two pre-existing units of high rank, the new unit may be included in one of the pre-existing units if its lithology falls within the pre-existing definition. Boundary stratotypes are still desirable, but their definition should be seen as a long term aim, to be stabilised only when a major project at a site has been completed.

(3) Isolated sediment bodies will normally have informal status of a type that clearly indicates the stratigraphic uncertainty involved. If possible, any <u>probable</u> stratigraphic position and <u>probable</u> membership of a higher ranking unit

should be indicated, with a clear presentation of the correlation methods used. Sometimes a whole site may consist of a series of equally important sediment bodies, isolated one from another, either by true gaps or by sediments which it is impractical to expose. In this case, initial lithologic description should be concentrated upon the definition of low ranking units (usually beds or lithozones, infra) in each sequence, followed by correlation of the different sequences. If it can be demonstrated that the correlation is likely to be reasonably accurate, higher ranking units may be defined, based upon the composite stratigraphy. Such a step is only justifiable if each stage is totally explicit, and especially only if each and every low ranking unit in each and every exposure has been properly defined beforehand. If this has not been done and an error is later discovered, the whole classification may collapse. Also, a lithostratigraphic classification is the basis for more advanced study and should not be reduced to a mere statement of general similarities. For instance, if two exposures show a similar division into upper calcareous deposits and lower siliceous deposits, it would be inadvisable, and of absolutely no utility, to define a 'calcareous member' and a 'siliceous member' covering the two exposures if no convincing correlation could be demonstrated between at least some of the individual beds of the two exposures.

(4) Correlation has been mentioned several times in the above discussion. A degree of lithologic correlation, often backed by lithogenetic arguments, is always necessary. It is useful if such correlation can be supported by biostratigraphic or chronostratigraphic arguments. In those cases in which lithology is rather homogeneous and several

different correlation schemes are possible, biostratigraphic and chronostratigraphic data may be used as the decisive factor. However, when there are no lithologic links between exposures which contain identical faunal or floral assemblages, or which clearly date from the same period, there is no reason whatsoever to merge the lithologic units into what would effectively become a biostratigraphy or a chronostratigraphy. Different lithostratigraphic units, of the same or different rank, may properly represent all or part of the same period, since sedimentary sequence has both a temporal and spatial aspect (cf. p.108). Correlation is discussed in greater detail in Chapter 17.

The concept of unit rank deserves some attention.

> Formations are the only formal lithostratigraphic units into which the stratigraphic column everywhere should be divided completely on the basis of lithology. (ibid., p.32)

Uncharacteristically for the Guide, this statement is equivocal. Nevertheless, whether the word 'completely' is part of the clause which follows it or not, the word 'everywhere' still suggests that all formal, and thus preformal, units must be defined as being part of a named formation. Those isolated deposits which cannot be convincingly correlated with the main sequence will remain informal units (unless they themselves merit definition as a separate formation); even within a generally preformal classification, there is no obligation to link such units with a formation. Some older stratigraphic codes have required that a formation be mappable at a given scale, eg. 1:25000. Obviously, most cave sites are not mappable at such small scales (a 100m length of cave would produce a map only 4mm long!). The Guide recommends no scale, but suggests:

> Practicability of mapping and of delineation on cross sections is an important consideration in the establishment of formations. (ibid., p.32)

As we have seen above, only those cave deposits which are accessible are mappable, and these will usually represent only a small proportion of the total volume. It is therefore suggested that the only way to apply the concept of a formation to cave sites is to leave the definition spatially 'open-ended', in accordance with note (1) on p.153.

The Guide recommends that there be no set range of thickness for a formation; it may be a metre or a thousand metres thick. This has led some researchers to suggest, in conversation with the author, that the entirety of the deposits within a cave, including those which are not yet known to exist, should be defined automatically as a formation. This is totally unacceptable to the author's mind, since the definition would not only have to be spatially 'open-ended' (which is unavoidable) but also compositionally 'open-ended'. The term 'formation' would become a convenient catch-all with no lithologic definition and thus no lithostratigraphic status or utility. There are plenty of informal terms already available for this theoretical entirety: deposits, sequence, succession, sediments, etc. The term 'formation' must be used in its proper sense, with at least a clear definition of lithology. It is mere prevarication to suggest that this criterion may be avoided by saying that a formation may still represent the entirety of the deposits because all deposits will theoretically have the unifying factor of material derived from the local bedrock. In practice, the presence of bedrock derivatives may indeed be the only obvious unifying factor for the components of a formation, but this does not remove the necessity, or desirability, of questioning whether

each newly discovered deposit may usefully be included in the defined formation. Occasions may arise where it is necessary to redefine the diagnostic characteristics of an existing formation, in order to exclude a new unit which is more usefully perceived as part of a separate formation.

> The degree of change in lithology required to justify the establishment of distinct formations (or other lithostratigraphic units) is not amenable to strict and uniform rules. It may vary with the complexity of the geology of a region and the detail needed to portray satisfactorily its rock framework and to work out its geologic history. (ibid., p.32)

This clearly places the onus on the utility of the distinction (cf. the similar arguments in Chapter 3). Nevertheless, since cave deposits are already rather small scale, it would seem undesirable to define a plethora of formations in a single site. It seems reasonable to suggest that boundaries of formations should usually be placed at major changes in either gross composition and/or overall fabric and structure, and only then if the distinction will be useful. One of the main reasons for recognising the usefulness of a distinction will be the possibility of <u>lithologic</u> correlation with other defined units (usually of at least formational rank) beyond the cave site, or of <u>lithogenetic</u> correlation with regional morphostratigraphic entities (cf. Chapter 17). The rank of 'member' will more usually be given to units that show important mutual differences within a site, but which are likely to be of only local lithologic significance. It has long been an established and useful practice in lithostratigraphy to allow considerations of areal extent and ease of correlation to influence the choice of status for higher ranking units.

We are left with the proposition that all cave/shelter deposits will contain at least one formation. This will be true of even tiny sites which we feel are sufficiently important to

treat at least on an explicit preformal level. The right to define our study material in this way must not be suppressed by ridicule from those who are used to dealing with much larger sediment bodies. The only concession that the present author believes reasonable is the systematic use of the explicit labels 'cave' or 'rockshelter' before the rank term (when the name of the formation would not naturally contain the word, eg. the ficticious Fairy Grotto cave formation or King George's Parlour rockshelter formation, but Henry's Cave formation). This will warn the reader of any synthetic text of a probable change in scale between cave and open-air formations.

The Guide suggests that geographic names should only be combined with the term 'formation' in a formal classification. However, the present author sees no pressing reason why this usage should not be extended to preformal classification. Note that, at least in Britain, the same name has often been used for two or more different cave sites; this is especially common when the name of an animal forms part of the cave name (the present author knows of eight 'Fox Holes'). Such sites will require a further geographic term to render the name of a formation unequivocal. It would also be useful if the discovery of new sites prompted the search for slightly more imaginative names. When two or more formations are present in the same site, each formation should bear the cave's name coupled with a descriptive term for the dominant lithologic characteristic (eg. the Fairy Grotto (cave) Scree formation, the Fairy Grotto (cave) Fine Sand formation and the Fairy Grotto (cave) Coarse Laminated Sand formation). The combined formations will usually have the rank of a group (eg. the Fairy Grotto cave group contains the above three formations), but this term must not become a catch-all

for all the observed and potential deposits of a site, any more than should the term 'formation'. The Guide also recommends geographic names for members (eg. the Fairy Grotto (cave) Shelly member). Traditionally, units within caves that have more or less member status have been numbered (eg. Unit III at the Fairy Grotto). The drawback with such a system is that it is not clear what a newly discovered unit, lying between previously named units, should be called. The drawback with a lithologically descriptive term is that, since the geographic term is invariable, one runs out of realistic names rather quickly and is often forced to use terms such as 'upper', 'middle' and 'lower' (eg. the Fairy Grotto (cave) Upper Shelly member and the Fairy Grotto (cave) Lower Shelly member, which may or may not be separated by a member with quite different lithology). Note that the use of the site name for both units and components of the same units (eg. the Fairy Grotto cave group and the Fairy Grotto (cave) Scree formation) is disapproved by the Guide, as is the use of terms such as 'upper' (and, presumably, numeric and alphabetic labels). The present author can see no way of avoiding such nomenclature, since it appears even more undesirable to divorce a unit from a named cave by varying the geographic term. It is therefore important that care be taken to avoid all possible misunderstanding (eg. the concept of the 'lower Upper Shelly member' should be clear, but that of the 'lower upper shelly member' is not).

The bed is the smallest formal (and preformal) lithostratigraphic unit. The desirability of achieving a correspondence between lithostratigraphic beds and lithogenetic beds has been noted above (p.137). A lithogenetic bed may, for certain purposes, contain several lithostratigraphic beds.

When it is clear that a lithostratigraphic unit must be defined that does not correspond to the traditional concept of a bed (i.e. it is not bedded, as in a zone of strong alteration that masks the original lithology), the term 'lithozone' may be used, although the Guide does not consider this to be a valid formal term. 'Lithozone' is also the recommended working term for all units of roughly bed rank; it may be used either pending more precise definition or because of stratigraphic uncertainty (eg. isolated sediment bodies). Individual beds and lithozones are not often named in geology in general but, in the cave context, they will all require some type of unequivocal label. Traditionally, beds are numbered which nearly always leads to problems with newly discovered units. Also, numbering encourages two common malpractices: renumbering to fit in new units, and illogical subdivisions (eg. layer 11a has nothing in common with layer 11 - it simply lies between layers 11 and 12). Whatever system is chosen, a unit name or label should only be changed when there is an overriding geological reason. The only system which the present author can think of that gives an idea of stratigraphic order, but which does not suffer when new units are found, is a multiple alphabetic system. Units could be given arbitrary 3- or 4-letter alphabetic labels, in stratigraphic order, with plenty of room 'in between' each label for new units. Such labels could also be combined with simple lithologic descriptions (eg. the MAC Scree bed and the MAT Sand bed are defined, and the MAJ Silt bed is later added). Such a system should work in most situations. Where bedding is highly lenticular and the exact ordering of minor units cannot be deduced, all attempts at ordered labelling must be abandoned.

When units, irrespective of possible rank, have been

demonstrably mixed, contorted or fragmented in such a way that the original internal sequence is obscured, they may be referred to as a 'complex'. Such situations are very common in caves. If several discrete pockets of obscure sediments are present at the same stratigraphic level, some attempt should be made to demonstrate their lithologic similarity before grouping them into a complex. Even in very patchy sediments, it may be possible to distiguish lithozones. Similar obscure sediments, that demonstrably occur at different stratigraphic levels, should not be grouped into a single complex.

A distinctive boundary between units that embodies important stratigraphic information which is additional to that provided by the bounding units (cf. p.141) may be termed a 'lithohorizon'. A lithohorizon may be time-transgressive, it is not necessarily horizontal and it will not necessarily remain confined between the same pair of units. A lithohorizon has intrinsic lithologic character and its extent is limited by the persistence of that character. A lithohorizon may be associated with either a thin, distinctive lithologic unit (a 'marker horizon') or with minor and discontinuous sediments (eg. the fill of bedding plane erosion features), as well as with mass post-depositional modification of underlying units (eg. a soil). In the absence of such features, the lithologic character of a lithohorizon is expressed in its structure and contact morphology alone (eg. a buried erosional bench). A certain amount of inference is often required to recognise the necessity for the definition of a lithohorizon. For instance, a soil may be placed stratigraphically as a lithohorizon, but a zone of even stronger alteration beneath a pool or stream deposit is best classed as a lithozone, since the boundary

itself does not supply significantly more stratigraphic information than the pool or stream deposit. The category of lithohorizon should be used sparingly and only when it is clearly useful; the definition must include objective criteria that will remain valid whether or not the inference concerning the cause of the phenomenon is correct.

There appears to be some disagreement in quaternary studies concerning the nomination of the statutory 'base unit' for lithostratigraphic purposes. Lüttig, Paepe, West and Zagwijn (1969) have proposed the bed, whilst Bowen (1978) has proposed the formation. To the present author's mind, this argument is something of a red-herring. It is perfectly reasonable that researchers interested in the greatest possible integrity of assemblages (geological, biological, archaeological, etc.) contained within lithostratigraphic units should stress the importance of beds. It is also perfectly reasonable that researchers interested in major changes (temporal or spatial) in sedimentary environment should stress the importance of formations. Bowen objects to the bed as a base unit because the "proposal fails to meet a cardinal requirement of a basic working unit, namely that it should be capable of being mapped as a tabular body of rock strata."(1978:87). As we have seen above, this objection would disqualify the great majority of caves from holding any sort of lithostratigraphic status at all. The present author cannot see why there should be a rigid base unit for all occasions. We may proceed by 'lumping' of beds to provide members and formations or by 'splitting' of formations to provide members and beds, whichever is the most useful approach in a given situation. However, it should be noted that stressing a specific rank as the base unit may have undesirable

effects. When one is concentrating upon the definition of a large unit like a formation, it is extremely easy to lapse into almost exclusively lithogenetic reasoning, with all the accompanying opportunities for errors of interpretation and correlation. The history of the stratigraphic study of glacial deposits in Britain illustrates this danger very well. More relevant to the cave context, however, are the dangers associated with the use of beds as base units.

West (1980) has provided a site-by-site description of the pre-glacial pleistocene deposits of the Norfolk and Suffolk coasts. Each lithostratigraphic unit at each site is defined informally as a "bed". The lithologic description for each "bed" is, from the present author's point of view, a little scanty, but it is usually sufficient for the purposes of West's study. The informal units are then grouped, or sometimes split, into formal "Beds", with no discussion or justification whatsoever. For instance, the Easton Bavents Clay (ranked equivocally as a Bed or Beds; ibid., Table 43, p.201) occurs at Easton Bavents, Covehithe, Aldeby and Ludham, that is, at four points stretched out in a rough line more than 44km long. The Easton Bavents Clay is then grouped with two other Beds, the Sidestrand Sands and the Cromer Stone Bed, to form the Sidestrand Member of the Norwich Crag Formation. The Sidestrand Member has a greatest linear dimension of over 72km, and there is a geographic gap between the exposures of the Easton Bavents Clay and those of the other Beds of some 25km. West states (ibid., p.104) that the Norwich Crag Formation was defined by Funnel and West (1977). In this earlier publication, the Norwich Crag Formation does not contain a 'Sidestrand Member'. However, the Norwich Crag Member does contain the Easton Bavents Clay, so that, in the absence of any

contradiction by West (1980) and in the absence of any proper lithologic definition, we may assume synonymy. The Norwich Crag Member also includes the Westleton Beds, so that the linear extent of the Norwich Crag / Sidestrand Member now becomes some 81km. Now, it is clear that the formal units, and the relationships between units, discussed above all have considerable utility in terms of biostratigraphy (pollen and mollusca). But what earthly use are they as a lithostratigraphic classification? Once the informal beds were recognised, ensuring the greatest possible integrity for biological assemblages, lithostratigraphic nomenclature was used to define something resembling more a composite biostratigraphy than a system of useful rock units.

In the cave context, the bed (or lithozone of similar rank) will usually be the working unit because we are normally most interested in the integrity of assemblages. A unit, which would be called a 'bed' on purely lithogenetic grounds, might conceivably be subdivided (using lithologic criteria) into several lithostratigraphic beds in order to provide greater stratigraphic definition for the study of such material as faunal assemblages or mineral suites. However, once this obligation has been met, there is no reason to define higher ranking units based upon anything but significant changes in lithology. We might use lithogenetic, biostratigraphic or chronostratigraphic data to help correlate discontinuous exposures of beds, and to group them into a submember or member, but we should not use such data alone to choose the rank of the higher unit, nor to correlate or group disparate sediments. Only when there is overwhelming evidence of a major erosional gap or unconformity should a unit boundary be placed at a point of relatively minor lithologic change (Hedberg 1976:39). If

there is no good lithologic reason to use intermediate ranks, a lithostratigraphy may be composed entirely of beds grouped into a single formation. When a cave sequence is defined as a single formation, it is because we cannot find any lithologic reason to subdivide it (i.e. it is homogeneous) and not because we wish to stress its radical lithologic difference from any other known unit. If one is worried that a lone formation has little lithologic significance in its own right, it is a salutary exercise to consider what sort of lithologic change would be necessary to provoke a reassessment, and ultimately to define a change in formation, were new deposits to be discovered (cf. the definition of the Pontnewydd Cave formation, p.1165).

The question of stratotypes in caves is complex and, at present, a generally depressing subject.

> The designation of a type section (unit-stratotype) is essential in the definition of a lithostratigraphic unit. (Hedberg 1976:37)

The present author agrees wholeheartedly with this proposition, even in the context of preformal classification. However, in reality it is not the designation of a stratotype which counts - it is the survival of the stratotype once designated. Many caves in Britain are scheduled Ancient Monuments or SSSI, but that has not stopped cavers digging trenches through the Pontnewydd sediments or 'adjustment' of the deposits at Kent's Cavern for the purposes of tourism. The conservation of stratotypes in unconsolidated sediment will never be easy, but it is impossible without long term commitment backed by adequate resources. At present, this commitment is coming from a few individual landowners or societies (e.g. the Pengelly Cave Research Centre, Buckfastleigh, Devon), a few local councils (e.g. the Creswell Caves project

set up by Nottinghamshire and Derbyshire County Councils), the National Trust and the Nature Conservancy Council. However, few if any of these groups have sufficient resources, and the co-ordination, which could only be supplied by an organisation such as the QRA, is lacking. The present author takes the pessimistic view that the only thing to do with a cave stratotype at present is to bury it as deep as is humanly possible.

The application of formal stratigraphic principles to cave sites is a relatively novel approach. The only instance of which the present author is aware is the appraisal of the Makapansgat Limeworks and Sterkfontein hominid sites (Transvaal) by Partridge (1975), using the procedure outlined by Bishop (1967). Although admirably detailed, Partridge's work is qualified as a "progress report" and it is not known whether this interesting document has actually been published. Clearly, the most useful procedures will only become apparent after many more cave stratigraphies have been subjected to rigorous, but hopefully not too rigid, classification.

6. QUANTIFICATION

6.1. Why Quantify?

There is only one valid reason for quantification, namely the desire to extract and communicate relevant information. Precision has nothing whatsoever to do with numbers. A numerical analysis may often help to structure an argument more exactly, but such an analysis may also effectively conceal errors of logic. A quantified argument is not inherently more precise or reliable than a qualitative statement. Furthermore, although a well presented numerical argument may sometimes allow the reader to evaluate the argument with less reliance upon the personal experience or opinion of the author, no argument is ever completely self-contained. What of those factors which the author did not mention at all because he felt that they were not relevant? Quantification may be fashionable at present, but it is only justifiable if a real increase in understanding or reliability may be expected to result.

> A criticism raised at proponents of quantification in the earth sciences is that they give numbers to things just for the sake of it. This must be avoided. (Till 1974:4)

A distinction must be made here between statistics and statistical tests. Statistics are merely numerical data which are assumed to convey relevant information. Some statistics are simple numerical statements of greater or lesser precision ('this cave is five metres wide'; 'this cave has a maximum

diameter of 5.01m at the deepest point to which sunlight has been observed to penetrate'; etc.). Other statistics are what may be called 'summary statistics' or 'statistical estimators', in that they purport to convey a general truth based upon a number of measurements. Point estimates (a median, a proportion, etc.) are simple summary statistics. More complex summary statistics include some idea of spread (total range, interquartile range, etc.). Intuitively, we feel that a point estimate is less reliable when it is unaccompanied by some statement concerning the spread of values. Note that, to derive one of the cited summary statistics, we have used mathematics, but only at a purely mechanical level ('arithmetic'). If we calculate mean and standard deviation, rather than median and range, we need more complex mathematics, and the texbooks will tell us that we also need to make certain assumptions about the real distribution of the values of the property we are trying to estimate. For this reason, such estimates are known as 'parametric statistics'.

When we see a statement such as 'the mean was 100 with a standard deviation of 10', we might suppose that the statement meant some such thing as 'there is a 68.3% chance that the true value lies within the range 90-110', that is, we might suppose that the statement gave some sort of mathematical judgement concerning the reliability of the estimate - and we would be very wrong. All we have been told is this: (1) the sum of the values of all the measurements, divided by the number of measurements, is 100; (2) the square root of the result of dividing the sum of the squares of the differences between the value of each measurement and 100 by the number, less one, of measurements, is 10. In other words, all that is involved is arithmetic and arithmetic does not make judgements. This extremely important

point is often overlooked. There are many statistical analyses that are mathematically so complex that a sophisticated computing facility is nearly always needed: curvilinear regression, factor analysis, discriminant analysis, constellation analysis, etc. Often the help of an expert mathematician will be required. Nevertheless, such analyses are still only arithmetic.

Let us return to the definition of 'statistics' - numerical data which are assumed to convey relevant information. This definition may be extended in the light of the intervening discussion. Statistics are numerical statements about some property of the study material in which we are interested, based upon certain assumptions concerning the link between the collected data and the property. Only the data and the assumptions are determinate and the mathematics (arithmetic) involved, no matter how complex, have nothing to do with the relevance of the data or the reliability of the assumptions. In the above example concerning the significance of a statement about a mean and standard deviation, any judgement on our part about whether there is indeed a 68.3% chance that the true mean lies within one standard deviation depends entirely upon the assumptions that the statement refers to a sufficiently large, random sample from a normal (or related) distribution. Clearly, what we need is some way of examining our assumptions.

Statistical tests are applied to numerical data in order to assess the accuracy of certain assumptions (hypotheses) about that data. However, there is one major drawback: statistical tests also involve assumptions about the data. In any statistical test there is a relationship between the assumptions involved in the test itself and its power of resolution; the more one may safely assume about the data, the more the test is capable of reacting

to fine distinctions. Obviously, the assumptions involved in the test itself must not interact with the hypotheses one is trying to assess. Thus, with respect to our example about mean and standard deviation, we would not test for a deviation from normality first, since all such tests require that the sample be random; the first test(s) must be for deviations from randomness as, happily, such tests do not involve an assumption of normality. Avoidance of interaction between a test and a problem often becomes more difficult. Again, in our example, we might be worried that a very small sample size would not permit tests for a deviation from randomness to detect a significant deviation. We would then need tests to decide whether such a small sample could conceal a significant deviation from randomness; these tests could not assume either randomness or normality. Such tests in fact exist, although they are extremely conservative (low power), that is, they will cause us to suspect that a sample is indeed too small unless it is considerably larger than is really necessary, a fact which becomes apparent only in retrospect, of course. This underlines the nature of most statistical testing, namely that a new situation should be approached with as few assumptions as possible; the power of the tests may only be increased when at least approximations have been derived which will support the assumptions inherent in more powerful tests.

The assumptions involved in statistics are of two basic kinds: unquantifiable and quantifiable. An assumption of normality, for example, is usually quantifiable because an exact numerical description of a normal distribution is available which, through probability theory, gives an exact method of allowing for random effects in the sample. Many assumptions in statistics

are unquantifiable, that is, they appear 'logical' but they cannot be stated in mathematical terms. For instance, there are many non-parametric coefficients of similarity which allow such statements as 'an object which is spherical, red and smooth is more similar to an object which is spherical, red and rough than it is to another object which is cuboid, green and rough'. Such a statement appears logical but it cannot be transformed into a mathematical relationship (we could transform it into a symbolic form but this does not constitute a mathematical relationship). When a statistic involves an unquantifiable assumption, no statistical test can be rigorously applied. Even the results of an 'arithmetically' complex method such as constellation analysis cannot be tested mathematically, because this analysis involves the unquantifiable assumption that certain differences and similarities between objects are significant. This does not necessarily mean that the summary descriptions provided by such statistics are unreliable. Everything depends upon our willingness to accept (and sometimes even upon our ability to identify) the assumptions involved. Care should be taken not to mistake statements such as 'optimal groups were thus formed' or 'the parameter X was thus maximised' for statements of statistical significance. Note that a statistical test involves an 'arithmetic procedure' and a 'mathematical logic'. It is the latter which gives the effect of a test; the former may be used on its own as a purely descriptive tool, based upon some logical, but unquantifiable, assumption. Indeed, it is acceptable to apply a statistical test to statistcs involving unquantifiable assumptions, as long as the assumptions are identified and it is clearly stated that the significance of the test result is entirely dependent upon the validity of the assumptions.

The great majority of numerical devices (whether in the original numerical form, or transformed into diagrams) which are traditionally used in cave sedimentology are summary statistics. They prove nothing; they merely illustrate, as clearly as possible, some property or relationship which the author believes to be significant. Such devices are extremely useful and will probably continue to be the dominant form of numerical analysis for some time to come. The derivation of relevant statistics is not always easy and there are many pitfalls, associated primarily with sampling theory and scales of measurement. There are also many statistical tests which may be applied quite easily to cave sediment data, as long as the underlying mathematical logic is relevant.

6.2. Sampling Theory

Archaeological sedimentologists often make no reference to sampling theory whatsoever. When some reference is made, it is usually comparatively brief:

- Un sachet de 1 à 2 kg. environ de sédiment est prélevé par le fouilleur lui-même, sous notre contrôle, dans chaque carré et pour chaque niveau, si mince soit-il. [...] Dans certaines zones choisies à l'avance, 25 à 30 kg. de sédiment sont prélevés dans chaque niveau pour l'étude des éléments grossiers et la construction d'un diagramme granulométrique global. (Miskovsky 1974:43)

  [Trans.: A bag of c.1-2kg of sediment is taken as a sample by the digger himself, under our control, in each square and for each level, no matter how thin. [...] From certain zones chosen in advance, 25-30kg of sediment are taken from each level for the study of the coarse elements and the construction of a diagram showing overall granulometry.]

- [...] ... even this relatively small size limit [10cm] imposes the necessity of taking relatively large samples, commonly 10 to 20 kg, in order to assure an adequate statistical sample of the larger fragments. In general practice, one samples a vertical column of sediments from the top to the bottom of the stratigraphic succession .... The sampling column is chosen so as to pass through as typical a section of sediments as possible ....

At the Abri Pataud we have been able to expand our sedimentological study to include several sampling columns scattered throughout the site .... The goal of such a sampling program was to verify the assumptions that a single sampling column is really representative of the entire site .... (Farrand 1975:31-2)

- Des échantillons pour analyses sédimentologiques ont été prélevés en plusieurs secteurs du gisement chaque fois que nous en avons eu la possibilité.
Cela n'a été réalisable que dans le cas de remplissages en cours de fouilles. Dans les gisements ayant fait l'objet de fouilles anciennes, protégés et classés pour la plupart, nous avons dû nous contenter d'un seul échantillonnage sur une même colonne verticale.
L'échantillon devant être représentatif du niveau qu'il doit caractériser, son volume a été conditionné par la texture du sédiment. C'est ainsi que pour des niveaux composés d'éboulis et dont les éléments ont fait l'objet d'analyses morphologiques, des échantillons de 20 à 30 kg. ont été prélevés. (Laville 1975:30)

[Trans.: Samples for sedimentological analysis have been taken, whenever possible, from several areas of the site.
This was only feasible in the case of deposits undergoing excavation.
In sites where excavation has taken place in the past, most of which are protected and scheduled, we have had to be content with a single set of samples from the same vertical column.
Since the sample should be representative of the level which it should characterise, its volume is conditioned by the texture of the sediment.
Thus, for levels composed of scree and whose elements have been subjected to morphological analyses, samples of 20-30kg have been taken.]

- Random sampling from the various levels of the Die Kelders 1 sediments was carried out by one of us (F.R.S.) during the four years of excavation. The sampling was supplemented by several visits to the cave during 1973. (Tankard & Schweitzer 1976:298)

More general texts treat the subject in greater detail.

All laboratory work relies heavily on sampling experiments as sources of information. It is impossible to analyse the whole of a stratum and a comparatively small portion has therefore to be used as a sample. The problem is to make this small amount representative of the whole .... Since those who may read the results of the analysis and use them in further interpretation of the site will, in general, not know the sampling procedure that was undertaken, it is the responsibility of the analyst to ensure that his sample is truly representative of the parent material, and that it was taken with care and with a specific problem in mind. A properly taken sample is selected from a population in such a way that statistical theory is applicable .... In addition, the size of the sample needs to be large enough to permit of satisfactory conclusions about the problem being studied.
The process of random sampling is often misunderstood, ...
A representative sample is a selection of individual elements from a population in the exact proportion in which they exist in the original population. This representative sample must be random. A random sample is taken to estimate a representative sample of a population of unknown composition. The larger the random sample the greater the probability that it resembles a truly representative sample. All laboratory treatment of archaeological sediments is based on the assumption that a true random sample is present, and if this is so then any one item, parameter or

> characteristic in the original population is as likely to be represented in the sample as any other, and the collection is free from conscious or unconscious bias. However, it is almost impossible to eliminate unconscious bias unless a standard procedure of sample selection is undertaken.
> When a sample is taken it is important to know the reason for sampling, since this affects both sample size and sampling plan. For example, if a series of calcareous ditch silts are being sampled for molluscan remains, closely spaced samples or a monolith are recommended, whereas samples taken to establish the sedimentary process responsible for the sequence can be taken from the individual layers at wider intervals. The sample size required for insect remains may be several hundred times the size required by a pollen analyst. (Shackley 1975:23-4)

With respect to the first four passages quoted above, Shackley's suggestion that authors will not generally give full details of their sampling procedures appears well founded. This is a deplorable situation. Shackley's advice on this point, though well intentioned, is totally mistaken. Let an author describe exactly why and how he sampled and his readers will soon indicate to him whether his results are likely to be 'representative' or not.

Four main ideas seem to emerge from the above passages. First, the concept of random sampling is glossed by our authors, either in a statement that they achieved it (but not how they did so) or in an explanation of why it was impossible in a given site. Shackley makes much of the concept but again does not tell us how. The second concept, allied to the first, concerns the choice of what is representative, either at the sample level or in the form of a 'typical section'. This in turn leads to the third idea, namely that vertical columns of samples should generally be taken through 'typical sections'. The fourth concept - that concerning the need to adapt sample size to the choice of parameters to be analysed and to the variance of these parameters - is very often expressed by cave sedimentologists in terms of how to handle the coarse fraction, since this problem is something of a nightmare in our field.

Our authors are clearly aware of the major factors involved in sampling theory. But have they really come to grips with these factors? What does 'typical' or 'representative' mean? Shackley's definition of the latter is a theoretical ideal. Are the choices of sample size correct and upon what criteria are they based? If the first sample is taken in a proposed column, how can the second sample, the position of which is more or less preconditioned by that of the first, be a random sample of the second layer? If a sediment shows lateral trends in certain parameters, will not random sampling, without some form of positional weighting, turn out to be quite misleading? With a sediment body from which we require several samples, when the first sample has been taken, do we replace it before taking the second, and, if not, can we, or need we, calculate the statistical effect?

Some of the questions posed here will turn out to be 'red herrings' but it is clear that the whole area of sampling theory deserves much more detailed attention. Anyone who feels that the mathematics and logic involved are too complex, and that the issue can be safely ducked, should also resign himself to the fact that what he has in his plastic bags is an unquantifiable assumption, not a reliable sediment sample. A sample is a statistical entity and there is no escaping it.

It will be convenient to start this discussion with the consideration of sample location, leaving aside for the moment the question of sample size. Our first priority must be a clear definition of the target population. Griffiths, in a general discussion of geological populations, proposes three categories:

> 1. The hypothetical population: that volume of rocks that occurred in some given volume of the earth's crust. This population remains a hypothetical population because there is generally some ambiguity

> concerning those rocks not now represented in the volume of crust defined; i.e., some sediments are believed to have been deposited and then removed, but they may never have been deposited. This material, not present, cannot be sampled; hence the entire population is not equally available to any proposed sampling scheme. Because the absent material is not represented in the samples, there is a possibility of bias.
> 2. The existent population: that volume of rocks existing at present in a defined volume of the earth's crust. This population may be defined unequivocally but is rarely accessible in its entirety for sampling. Hence, while this is the target population for most geological investigations, it cannot be random-sampled; i.e., all parts of the population may be present but are not equally likely to be contained in the sample program.
> 3. The available population: the population represented by the accessible outcrops, road cuts, boreholes, etc. This population should be sampled properly, and conclusions from the samples may be applied to this, the available population, if the sampling program is suitably desiged. (1967:14-5)

It is easy to see how these categories apply to the cave situation. A fourth category may be added to the list:

4. The optional population: that body of sediment chosen for study by the sedimentologist from the available population. This population should be sampled properly, and the conclusions from the samples may be applied to this, the optional population, if the sampling programme is suitably designed.

The reason for defining a fourth category will be made clear below. For the moment, note that this category has perfect statistical validity. There is absolutely no trickery involved in sampling from such an apparently arbitrary population.

The best way to discuss the topic of sample location is to approach it by way of the existent population, the one claimed by Griffiths as being the target population for most geological investigations. First, let us assume that the existent population at a given site is also the available population. If we then define our optional population as the whole available population, we will derive statistics that are applicable to the whole site, such as mean particle size or carbonate content. But these statistics will not be applicable to individual strata, since one

of these was not the target population. Knowledge of mean parameters for a whole site will very rarely be of use to us. We may now subdivide the existent site population into strata, or groups of similar strata (eg. laminations within a bed), and refer to them as existent stratum populations. If we again define the whole of an existent, and in this case available, stratum population as our optional (stratum) population, we will derive statistics that are applicable to the whole stratum. But, as with the entire site, how often will we wish to have information concerning a whole stratum? Consider an influent stream deposit, that is very coarse near the entrance but which becomes progressively finer into the cave. Of what use is a mean particle size for this stratum? Similarly, in a low energy deposit, limestone scree will probably become less abundant into the cave.

In previous chapters, it has been suggested that cave sediments show extreme lateral variability. This variability is not random; in a statistical sense, the variability must not be confused with variance (although the sediments will also have a degree of variance). The variability is structured as a geographical or spatial trend, which may be more or less linear and continuous. In the cave situation, the rate of change in such a trend may often be very fast compared with many geological situations in higher and more constant energy regimes. The actual nature of the trend will often be the most interesting aspect of a sediment from our point of view. Our target population in this case is not a stratum itself, but the rate and nature of change within this stratum. If, however, we do wish to extract 'average' parameters for a whole stratum, the effect of the trend must be suppressed. We now wish to set up a sampling plan which, according to theory, must be random.

Consider what a random number generator really does (cf. also p.183). It produces a rectangular probability density function (abbreviated to 'rectangular p.d.f.'; cf. Wetherill 1972:66-7), that is, within set limits, any possible outcome (number) has the same probability as any other possible outcome. Now, it is obvious that, in trend analysis, a single sediment sample cannot give us any idea at all concerning the nature of the trend. Sediment samples must therefore be considered as individual observations (elements) in a sample of several such observations. Shackley (supra) rightly points out that (other things being equal) the larger the sample, the greater is the probability that it resembles a truly representative sample. If we take very large numbers of random sediment samples along a line following the axis or direction of the trend under investigation, we will find that the number of sediment samples within a given stretch of the line approximates the number of samples in any other stretch of the same length as the first. High-sample-number random sampling has achieved two ends: (1) samples are uniformly distributed along the trend (a rectangular p.d.f.) and (2) certain types of bias (infra) have been avoided.

Let us consider a hypothetical case in which we have nine metres length of section and we wish to examine a stratum which shows a clear lateral trend in the parameter of interest. We require an 'average' value for this parameter, that is, we wish to suppress the trend. We will take ten random numbers from a table of such numbers in order to define our sampling positions. We are aware that ten samples are probably not enough to approximate the 'very large number of samples' mentioned above, but, after all, ten is more than most people would take and, in any case, we cannot afford to take more at this time. It must be

admitted that the following outcome is possible:

(A) 0.00, 1.00, 2.00, 3.00, 4.00, 5.00, 6.00, 7.00, 8.00, 9.00

However, we decide intuitively that this sample is not in fact random. We might suspect that random sampling does not necessarily produce random samples. We may nevertheless expect such a sample to be a very rare occurrence. To be on the safe side, we will test all future samples for randomness. We take another sample of random numbers:

(B) 4.49, 8.98, 7.91, 3.94, 8.83, 7.73, 8.91, 4.40, 8.54, 8.63

This sample is much more haphazard than the first but the textbooks tell us that being haphazard does not guarantee randomness. A test is required.

> Experience with tests of random number sequences indicates that the runs test is probably the most discriminating of the tests listed here [Kolmogorov-Smirnov test, serial test, poker test, permutation test] .... (Maisel & Gnugnoli 1972:138)

We will use the tables calculated by Swed and Eisenhart (1943) for the runs test at a 95% significance level. Sample (B) shows no significant departure from randomness. Just as a comparison, let us also collect the ten samples indicated by our first set of 'random' numbers (A), even though we suspect that they are not in fact random.

Unbeknown to us for the moment, the parameter we are studying behaves in the following manner:

$$y = x$$

Where y = value of the parameter

and x = distance in metres from point 0.

We now measure the parameter in each of the ten samples, in each of the two sets. Because we are using a superb measuring technique, the answers are exactly the same as the sampling location numbers. We require an 'average' value for the parameter,

so we will calculate mean and standard deviation for each set:

(B) tested random sample : 7.24 ± 2.09

(A) probably non-random sample : 4.5 ± 3.03

It is clear from the equation $y = x$ that the true mean is 4.5. Why has the tested random sample failed so badly? And why has the other sample provided such a good estimate? Let us test (B) again for randomness, this time using the Kolmogorov--Smirnov test. This results in a departure from randomness which is significant at the 99% level. So it was the runs test which was at fault, thus explaining why (B) failed to give the true mean. However, and here is the apparent paradox, sample (A), which fails the runs test miserably, shows no deviation from randomness at the 99% significance level using the Kolmogorov--Smirnov test! Is there really something wrong with our tests for randomness? Could it be that sample (A) is, in some peculiar way, random?

This long hypothetical example was necessary in order to introduce the following idea:

> I contend that there is no such thing as absolute randomness .... In any sampling inquiry it is necessary to ask oneself, Is the sampling method I am using random for the universe [population] I am considering, for the characteristics I am discussing, and for the sampling distribution or tests of significance I am employing?
> Randomness is relative.
> (Kendall 1941:5)

Kendall's arguments in support of this suggestion are compelling and this article is accessible, granted only a very basic knowledge of formal logic.

If we return to the hypothetical example, we can see what really happened. First, neither of the tests used was at fault, since neither claims to be a test of randomness as such. The runs test identifies patterning in a sequence of numbers. Thus, sample (B) which was haphazard was not rejected, whilst sample (A),

which was highly patterned, failed the test. The Kolmogorov--Smirnov test identifies departures from a specified frequency distribution. Truly random numbers have a rectangular p.d.f. Although it was not haphazard, sample (A) had a rectangular p.d.f., so that it did not fail the test. The small sample size of sample (B) meant that its distribution did not sufficiently approximate the theoretical rectangular p.d.f. and it therefore failed. A 'true' test for randomness was not used because no such test exists. It is clear that the sort of randomness required to give a good estimate of the mean in our hypothetical example was a good approximation to a rectangular p.d.f. A sample generated by random numbers could have achieved this aim but a much larger sample would have been required. The presence of patterning in sample (A) was totally irrelevant to our problem.

Because the concept of randomness is so important to sampling theory, it is worth trying to see why there are so many pitfalls in its application. According to Snedecor and Cochran:

> [...] ... random sampling means that every member of the population has an equal chance of appearing in the sample, independently of the other members that happen to fall in the sample. (1967:10)

This definition is unequivocal. There are two parts to the definition: (1) equal probability for each member, giving a rectangular p.d.f., and (2) the choice of one member must not influence the chances of selection of any other member. Now try to devise a method of collecting a random sample. No such method exists because there is no way in which the two parts of the definition can be combined into an exact mathematical expression. In practice, pseudo-random numbers are generated that give the best possible approximation to randomness, but which eventually fail both conditions.

Most modern pseudo-random number generators use a

'congruential' method. This involves the choice of four non-negative integers: X_0 is the starting value (the 'seed') for the generator; A is called the 'multiplier'; C is called the 'increment'; and M is the 'modulus'. M must be greater than both A and C. The number is given by the following formula:

$$X_{n+1} = AX_n + C \ (\underline{\text{mod}}\ M)$$

Starting with X_0, we multiply it by A, add C and divide by M. We then take the remainder, losing the decimal point, as our pseudo-random number and also the next 'seed' (eg. if $M = 10$ and $AX_n + C = 32$, then $X_{n+1} = 2$, etc.). Clearly, we have already departed from the second requirement of the definition of a random sample, since any number in a sequence is determined by the preceding one, albeit via an obscure transformation. The choice of the values for X_0, A, C and M are very important. Try the values 3, 5, 4 and 10; the resulting sequence '399999...' does not look very random and it does not have a rectangular p.d.f. With 0, 1, 1 and 10, we get '0123456789012 34...'. All congruential generators have a period (which at best cannot be greater than M) at which they repeat. The period will be equal to M if (1) A-1 is a multiple of every prime number dividing M, (2) A-1 is a multiple of 4, if M is a multiple of 4, and (3) the only divisor common to both C and M is 1. Obviously, we would not like the look of a repeating sequence, so we would make M larger than our sample size. But if we do this, we will be adding the restriction that any number may only occur once and that some possible outcomes will not occur at all. In fact, there are rules, derived from Number Theory, which help to overcome many problems with pseudo-random number generators (cf. Knuth 1969).

Although 'packaged' pseudo-random numbers are readily available, the above discussion is far from trivial. There are

situations in sedimentology in which thousands (or even millions) of random numbers might be needed and it would be too laborious to get them from a reliable source such as published tables. The RND functions on most 8-bit microprocessors have periods of only one or two thousand - sufficiently long, however, for it to be extremely difficult for the operator to spot the repetition, since the cycle does not necessarily start at the first number (i.e. there may be a string of a few hundred numbers, followed by a second string of a few hundred; the second string cycles but the first is never repeated). Even main-frame generators usually repeat at only a few thousand numbers, in order to economise on computing time. It is no good just programming a 'good' algorithm referenced in the literature, since generators interact with the way in which numbers are stored and rounded in different computers. When using a pseudo-random number generator, it is important to ask the staff about the period and whether any tests for different sorts of deviation from randomness have been carried out on <u>that</u> generator on <u>that</u> model of computer. There is no way to construct a good software pseudo-random number generator (a mathematical algorithm) for a microprocessor system. The only solution here is to interface a hardware generator (eg. using the electronic 'value' of a noisy component such as a transistor) but, even with a non-periodic generator like this, care must be taken to ensure that there is no time-dependent departure from randomness (eg. caused by gradual heating of the component, etc.). The difficulty of devising a good generator (or of finding an existing system which actually lives up to its reputation) has, so far, prevented the present author from making very much progress in a project designed to approach the thorny problem of particle size analysis in caves from the direction of simulation.

In summary, we may say that randomness may only be mathematically defined by what it is not. Since there are an infinite number of things randomness is not, the only possible approach to an application of the concept is to determine the most important properties that the data should not have, given the problem in hand, and then to test for these (cf. Tocher 1963 for further discussion).

This brings us to a consideration of the concept of bias.

> Any estimate whose average value over all possible random samples is equal to the population parameter being estimated is called unbiased. (Snedecor & Cochran 1967:45)

This definition of bias is both minimal and sufficient. Thus, the estimate of the mean given above using sample (B) was unbiased, even though it was far from the true mean. This is so because the average of the means of every theoretically possible random sample of 10 would be 4.5. It is sometimes said that randomness and lack of bias may be equated:

> This sampling procedure [random sampling] is said to be unbiased. (Till 1974:48)

This is not strictly true but, as long as the true nature of bias is understood, no harm can result.

Let us return to the accurate definition of bias given by Snedecor and Cochran. It would appear that we must take all possible random samples and have a knowledge of the true population parameters in order to decide whether an estimate is biased or not. In effect this is true, but the calculations are done on the theoretical not the empirical level. We must understand the nature of the population parameter (or be able to approximate that nature) in order to test whether, in theory, all possible random samples (of the sort we are actually going to take) would

give a correct estimate. The most common example of the analysis of bias takes the normal distribution as the model for the distribution of values of the parameter in successive samples from the population. Other distributions may, of course, be specified. Bias is therefore a feature which must be calculated, and to do this we must have random sampling and a knowledge of the population. If we cannot fulfil both these requirements, the concept of bias is meaningless. That is, we cannot say that the estimate derived from sample (A) above is unbiased. This does not imply that this estimate, or others like it (eg. the median of any sample), are biased - the concept of bias is simply not applicable.

Having said this, it should be pointed out that the word 'bias' is often used to cover all sorts of systematic errors. When statistical analysis is being discussed, the restricted usage given here is preferable since, if we accept the wider usage, we can no longer say that random sampling as a procedure is unbiased. 'Biased' would merely take on the common meaning of 'misleading' and we could never define it exactly.

If we feel justified in using the concept of bias in an analysis, we can employ it to gain some idea of the magnitude of other systematic errors. In the case where we feel a normal model is adequate, the analysis of variance should highlight some other types of error that might be affecting our results. If the normal model cannot be assumed, but we nevertheless feel that the mean is still a valid parameter, the Central Limits Theorem may be used. Note that such analyses are not guaranteed to register all types of error.

In the above discussion, we have introduced two types of sampling plan, the simple random sampling plan (cf. sample (B)) and a type of highly patterned plan (cf. sample (A)), referred to

as a systematic sampling plan. There are also several other types of plan, such as the stratified sampling plan, which are useful in some sedimentological analyses, but these will not be discussed here. A full treatment of sampling is given by Cochran (1977).

Systematic sampling is particularly useful in the presence of a lateral trend. If the trend is constant and linear, or can be thus approximated by an algebraic transformation (eg. by using log-distance rather than distance), the resulting rectangular p.d.f. is less likely to cause error with all forms of linear regression analysis than a random sample of the same size (cf. Snedecor & Cochran 1967). Sampling positions can easily be calculated, using the excavation grid, probably already available in the cave, and a single random number to define the starting point. This last detail removes subjectivity (so-called 'conscious bias'). If the trend appears to be accelerating or decelerating, the distance beween samples can be varied, say by doubling or halving the interval, without introducing new types of error into the plan. It is very difficult to do this with a random sampling plan. If a trend is irregular (changes of pace or even direction), a traverse can be split up into sections where the trend is roughly linear, again without having to start a completely new plan. If an 'average' value of a parameter is required in the presence of a trend, systematic sampling is still considerably more efficient than random sampling (cf. Cox 1958; Cochran 1977), since the trend effect can be suppressed by weights derived from the trend analysis. Note how, in the hypothetical example given above, sample (A) was self-weighting (because of the equation $y = x$), producing an exact estimate of the true mean.

When using a systematic plan, two factors must be kept in mind. Although there are some very approximate methods of

applying the concept of bias to systematic sampling, it is still true, as was suggested above, that "no trustworthy method of estimating $V(\bar{y}_{sy})$ [the variance] from the sample data is known" (Cochran 1977:231). Without the basis of randomness, we cannot calculate how much our estimate is likely to be in error. This would normally preclude the use of systematic sampling altogether were it not for the fact that it is so much more efficient than random sampling. Although we cannot estimate error from the original data, we can test the performance of the trend analysis by taking a small number of samples at random positions. The second drawback of systematic plans is that they are subject to truly massive error if there is some harmonic effect between the sample interval and periodicity in the parameter under investigation. Consider how inaccurate an estimate of the average proportion of some exotic mineral might be if the samples fell systematically under the regular fissures in the rock through which the mineral was entering the cave. It is not usually very difficult to avoid such error, just as long as the problem is thought out before the samples are taken.

Random sampling should be used whenever there are grounds to suspect that no lateral trends are present. Such a plan is often useful on comparatively small exposures, where any trend will have little effect on the results. The obvious advantages of random sampling lie in the fact that the results can be checked for possible error and a whole range of parametric and non-parametric statistical tests can be applied.

The last type of sampling plan that should be mentioned here is also the most commonly used: subjective choice by the sedimentologist of what he considers to be 'typical'. Such a plan should not be ridiculed; given the constraints of cave

sedimentology, it is often the only plan available. This brings us back to the definition of the four different categories of population given at the beginning of this discussion. As sampling locations become more restricted, the target population becomes similarly restricted and the validity of statistical inference becomes more localised. Put in this manner, it is clear that, no matter how restricted a sampled exposure, statistical analysis is valid as long as no spurious generalisations beyond the target population are claimed. There is no point in complaining about the discontinuous, non-random and generally unsatisfactory exposures that are usually all that is available to us, unless it is actually within our current capabilities to extend those exposures. We can achieve much by deciding whether, within the available population, we may define optional populations that are worth sampling. The definition of the target population is always a subjective assessment on the part of the sedimentologist; its validity depends entirely upon the quality of the arguments used in its construction. There are no simple mathematical aids towards definition and a blind trust in complex random sampling plans is no substitute whatsoever. There are many cases in which the definition of an extremely restricted optional population will be of use to us. Sampling from a column of sediment with an exposure width of only a metre may be valid if we require information concerning, say, exotic sand grains entering the cave through time via a roof fissure directly above the column. We may even define a single sediment 'sample' as an optional population in its own right. Every decision concerning sampling plans should be accompanied by a clear definition of the optional population, by step by step reasoning from the particular to the more general, and by description of the measures taken to avoid

systematic error or bias.

There is much confusion in the literature concerning the reason for taking vertical columns of samples. In the example given above of exotic grains entering via a roof fissure, it is clear why such a sampling plan is relevant. However, it is often implied that vertical columns always ensure, in some mysterious manner, against systematic error. This is untrue. It is often useful to hold the horizontal positions of samples constant in order to examine the differences between strata that are thought to be primarily due to a process acting at a particular location through time, or in order to examine the lateral 'migration' of a process through time. However, if this position is held constant merely on the assumption that samples from different strata will always be in some way comparable, simply because the samples are located one above the other, then the plan is 'neutral', that is, the effect of the plan is unknown. It may be argued that a vertical plan ensures that samples are in the correct stratigraphic order, that is, there is no temporal overlap or inversion. This is often, but not always, true, and this may be a useful principle for work within a single stratum. However, the author would have thought that, if an operator is such a bad stratigrapher as to have to resort to such mechanistic rules for a whole sequence, he should not be sampling in the first place, since he is tacitly admitting that he cannot recognise the relevant boundaries and bedding features. The only case in which a strictly vertical plan should automatically be used (at least as a first approximation) is when mobile material (eg. pollen, interstitial precipitates, etc.) is to be examined.

Before moving on to a discussion of sample size, we should consider the allied question of sample number, that is, the size

of a sample of samples.

> [...] ... the number of samples chosen should be proportional to the magnitude of the variability. (Griffiths 1967:22)

Variability may be structured, and thus constitute a trend or a periodicity, or it may be less obviously structured and best referred to as 'heterogeneity' (akin to the statistical concept of variance). It is clear that, in this case as with all statistical problems, the more we know about the population to be sampled, the easier it will be to design an efficient analysis. In some cases, we will know so little about the population that a pilot project will be necessary to estimate the magnitude of the variability. This iterative method has rarely been used in cave sedimentology, despite its power and efficiency.

Everything that has been said above about sample number is also applicable to sample size. Let us try to put some of these ideas numerically.

As a first example, let us consider a case in which we require an estimate of the mean of some parameter. We will use the Central Limits Theorem, which states that, given random sampling and irrespective of the original distribution of the parameter values in the population, the frequency distribution of sample means will approximate to normality as the number of samples increases. We need to decide how much error we are willing to accept in our results. Suppose that this error is L and we will allow a 5% chance that the error will be greater than L. The Central Limits Theorem leads us to expect that our means will be normally distributed with 95% confidence limits on the mean of means, $\bar{X}$, as follows:

$$\bar{X} \pm 1.96\sigma/\sqrt{n}$$

Therefore, our acceptable error, L, must not exceed this

value:

$$L <= 1.96\sigma/\sqrt{n}$$

The sample size, n, required to satisfy our stipulated case is:

$$n = 3.8416\sigma^2/L^2$$

To calculate n, we need to decide upon the maximum acceptable error and the acceptable chance that the error will in fact be greater. We also need to know, or to derive an acceptable estimate of, σ^2, the variance. It is often difficult to estimate σ^2 if no previous work has been done on the sediment in question, or on similar sediments (supra). If the parameter is unlikely to be normally distributed, it may have a rectangular p.d.f.; in this case, Edwards Deming (1960) has suggested that if:

h = highest value of parameter minus lowest value

then,

$$\sigma^2 \doteq (0.29h)^2$$

As a second example, let us consider a case in which we require an estimate of a binomial proportion, p. As before, we decide upon L and again we accept a 5% chance that the error will be greater than L. The formula now becomes:

$$n = 3.8416pq/L^2$$

where $q = (100-p)$

and p, q and L are in percentages.

Here, we need an estimate of p in the population. Note how the product pq is small if either p or q is small. If p and q are nearly equal, pq is large, and so is n.

We now turn to the most intractable problem in sampling theory as applied to cave sediments, namely, the sample size required to achieve good estimates of the proportions of coarse particles. The parameter, n, expresses sample size as the number

of elements included in a sample.

> [In particle size analysis] the sample size is artificial, a percent by weight; in such a case with sample size n unspecified, the degrees-of--freedom concept is difficult to evaluate, and so tests of significance are impossible to perform. The 'statistical estimators' obtained are therefore obscure. (Griffiths 1967:92)

> [The degrees-of-freedom concept] reflects the number of observations that are free to vary after certain restrictions have been placed on the data. (Siegel 1956:44)

> In general the degrees of freedom are the number of observations minus the number of estimates made from them. (Till 1974:56)

The 'statistical estimators' to which Griffiths refers are such things as the Inclusive Graphic Statistics of Folk and Ward (1957) which are widely used to describe particle size distributions.

To apply statistical tests or to estimate the required sample size, we must have an estimate of the parameter n, the number of particles. There are only four options open to us. First, we can count the particles, despite the labour involved. This option is very rarely practical in the cave situation, although some finer sediments may be treated by an indirect method in a Coulter Counter (cf. Behrens 1978). Second, we may ignore the individual particles and take a large 'sample of samples'. This is the option most commonly used in the geological sciences in general (cf. Griffiths 1971), but it is dependent upon massive resources (i.e. commercial applications) which are unlikely to be available to many cave researchers. The third option is to ignore statistical significance and numbers of particles, and to base the decision as to particle size, expressed as a volume or weight, on experience and especially upon empirical findings concerning reproducibility. This is the option used by most cave researchers, although it is not usually clear whether they have conducted any formal reproducibility checks. Many other specialists, such as soil engineers or pedologists, also use this method, which is formalised in national and international codes of practice

(e.g. British Standard 812, 1951). It is very important to understand that such codes take two factors into account: rough estimates of the sample size needed with respect to the type (particle size distribution) of sediment involved and the tolerances of specific measurement techniques, often themselves described as standards by the same national or international organisation. Thus, Allen (1975) gives recommended sample sizes for dry sieving analyses, in which the sample sizes are defined more by the tolerance and performance of British Standard sieves than by any clear idea of sediment-related significance. Measurement-related factors cannot, of course, be ignored in any analysis, but it is nevertheless imperative that careful thought be given to sediment-related factors. Having taken as many factors into account as possible, and assuming direct measurement of individual particles of diameters of over 8mm, the present author has made the following very general suggestion:

> On theoretical grounds, the writer would recommend that a 2kg sample be taken for particles up to 20mm diameter, and that 1kg be added for every increase of 1mm; thus, 32kg would be needed for samples containing 50mm particles, 82kg (sic) for 100mm particles. This may seem ridiculous, but if anything, it is likely to be an underestimation of the true weight needed .... (Collcutt 1979a:292)

For the time being, the author stands by this statement, although reproducibility checks have shown that, in those cases where much of the sediment (>c.25%) lies between 1 and 20mm (i.e. 'gritty' deposits), it is best to increase the sample weight to at least 5kg.

Three out of the four options have now been discussed. The fourth option involves attempts to estimate the number of particles by theoretical means. If the particle size distribution can be described (or reasonably closely approximated) by a precise mathematical formula, we will have the first element needed for

an estimate of required sample size in terms of numbers. The most common case arises when the weight distribution is log-normal, a situation which is often approximated in certain high energy sediments (cf. p.215) and which may be defined exactly in theory. To the author's knowledge, the mathematical link between weight (or volume) frequency and number frequency, when the former distribution is log-normal and particles are idealised as spheres, has never been published. Proofs are therefore offered in an Appendix (p.1231) to the present text, both in a long and purely mathematical form and in a shorter statistical form (requiring considerable knowledge of statistical theorems). Not only is this link immediately useful in computer simulation studies, but it also provides an exact figure for the number of particles in a perfect sample (one that has an identical particle size distribution to the population) of given weight (the convertion from volume to weight is made simply by selecting an 'average' particle density). Note that the number and weight distributions have been proven to be of identical shape (both are log-normal and they have the same standard deviation), the mean of the former simply being much smaller (larger in terms of <u>phi</u> units) than that of the latter. The next step, now that we can actually define a number population, is to derive a formula for confidence intervals of proportions within certain size limits. With the help of M.J. Collcutt (a mathematician) and Professor D.E. Barton (a statistician; Queen Mary College, London University), the author is pursuing this topic. An approximate formula has already been derived, but there still seem to be considerable difficulties, probably linked with a significant window effect in this complex PPS (probability proportional to size) sampling problem. As a rough analogy, the situation is as follows. If one asks a man in the street how

many siblings he has, and he answers "three", one can record one quarter of a sibling group of four, without having to go and find the other three siblings and include them physically in the sample. However, when taking a sediment sample, one cannot record the presence of one quarter of a large particle simply because one quarter of it sticks into the defined sample volume; one must take the whole particle in order to recognise its nature but, by so doing, the sample volume (weight) is increased and the volume and number distributions are skewed. This difficulty is not insurmountable but it will require considerable additional work. Note that, although this sampling problem has been approached using the log-normal distribution and spherical 'particles', it could theoretically be solved (with a good deal of mathematics) for any defined particle size distribution and any defined particle shape or shapes. Once the format of a solution for the simplest (!) case has become available, it will be much easier to derive formulae for more complex (and more realistic) situations.

As a conclusion to this section, the reader might like to consider the following proposition:

> This, then, is the problem of sampling: how does one interpret the answer? (Griffiths 1971:B347)

6.3. Numbers, Scales and Indices

There are four main types of scales which can be used to measure a property:

(1) A nominal scale uses numbers only as arbitrary symbols to indicate difference. The property being measured is divided into a number of discrete 'attributes' or 'attribute states' but there is no implied numerical relationship (i.e. no scale of

difference or order of 'attributes') between the different 'attributes' (e.g. numerical coding of rock types).

(2) An ordinal scale uses numbers to indicate an intrinsic ordering of values, without implying a scale of difference (e.g. the residual facies discussed in section 3.3.).

(3) An interval scale uses numbers to indicate both an order and an exact relationship (usually arithmetic, but also geometric) between successive values (e.g. degrees centigrade).

(4) A ratio scale uses numbers to indicate an order and a relationship between values; in addition, the scale involves a true zero (e.g. dimensions such as length).

Interval and ratio scales are together known as numeric scales, since values may usually be manipulated by the normal rules of arithmetic. Nominal and ordinal values must never be treated as numbers subject to ordinary arithmetic. Discrete values are always taken by both nominal (the 'attribute') and ordinal ('greater than', 'equal to', 'less than') scales. Numeric scales may take either discrete values (e.g. counts) or continuous values (e.g. dimensions), although all values are in fact discrete in as much as a value is quoted only to a certain number of decimal points. Measurements on a given scale may be transformed in various ways. A continuous scale may be divided into equal classes (or classes defined by a regular mathematical relationship) to give a discrete distribution, or into classes with any intervals to give an ordinal scale. Such transformations always reduce the 'power' (information value) of the measurements, but they may be useful, either to simplify the measurement process itself (e.g. particle size distribution is continuous, but it is usually easier to take classed measurements), or because there is doubt as to the relevance of the initial scale (e.g. the thickness of beds

is measured on a ratio scale, but it may be desirable to reduce the measurements to an ordinal scale before examining an unknown relationship between bed thickness and sediment type). Ordinal scales may also be used to measure conceptually continuous variables for which there is no simple measurement technique (e.g. subjective sorting of stones into roundness classes, infra). A 'weak' scale should never be transformed into a more 'powerful' one.

A data batch may sometimes be described by various summary statistics. Nominal data may be described by the mode, the most commonly occurring attribute or attribute state. Intrinsically ordinal data may not be described by any summary statistics at all (e.g. the 21 layers of a site are in an intrinsic order, but the tenth layer is not necessarily 'average' in any way). However, most ordinal data in fact represent conceptually continuous variables and may therefore be cautiously described by the median, the nature of the individual holding the central place in the ranking. Similarly, the range has meaning for ordinal data only if they have been transformed from numeric data (e.g. the range of bed thicknesses may be given, even if thicknesses are then to be transformed to an ordinal scale for some reason). Numeric data may be described by a literally infinite series of summary statistics (including mode, median, mean, range, variance, etc.).

It was said above that nominal data may be described by the mode. It therefore follows that all the attributes could be ranked in order of frequency. This does not mean that the measurements are now on an ordinal scale, merely that the totally discrete nature of the attributes allows their occurrence to be counted on a numeric scale. This in turn allows the application of powerful statistical tests, based upon discrete theoretical

distributions such as the binomial. Caution is called for if counts are derived from the reduction of a continuous scale to a set of discrete classes; it may be necessary to include a correction for continuity in order to apply tests derived from discrete theoretical distributions to such counts.

Counts on their own, although on a ratio scale, have little meaning. Counts become useful when they express a relationship between two or more sets of data, or between data and a conceptual standard. It is often convenient to standardise counts by reducing them to percentages or proportions. By so doing, two limits are set to the new scale. Some properties are nearly always presented in this way: rather than saying that a surface drops 11.3cm for every 174.8cm, we say that the slope angle is 3.7°. Instead of setting limits on possible values, continuous variables may be standardised by pulling their central (mean) value to zero and reducing their variance to unity:

$$z = x - \bar{x} \,/\, s$$

where z = each new standardised value

x = each measured value

$\bar{x}$ = original mean

s = original standard deviation.

This is a useful device for comparing skewness or peakedness of two distributions that originally had different means and spreads of values. Although standardised measures are still on a numeric (but now interval rather than ratio) scale, they are essentially dimensionless. There are special rules for the manipulation of such values and they should not be treated as ordinary numbers. For instance, to say that three samples, with 40%, 50% and 60% of a given constituent, have an average of 50% of that constituent will <u>not</u> necessarily be true (meaningful).

There are other kinds of standardised measures that may be referred to as indices. An index is essentially the same type of number as a proportion or an angle. The difference lies in the fact that an index is defined for a specific purpose and has absolutely no meaning beyond that purpose. Thus, a proportion of 0.1 is a simple standardised measure, whilst a statement such as 'breadth/length is 0.1' may be considered to refer to an index. Indices are not restricted to cases in which there is an obvious numerical relationship between the two or more terms of the calculation. A 'breadth/length' index has two terms measured on the same ratio scale, but it is perfectly valid, and often useful, to define a discretional index such as 'weight/length' or 'carbonate content/bed thickness'. Note that the scale of a simple index is still explicit (a 'breadth/length' index of 0.1 means that the object is ten times longer than it is broad), but that a discretional index is not (a 'carbonate content/bed thickness' index of 0.1 has no intrinsic meaning beyond the arbitrary algebraic definition).

Sedimentology employs all of the types of numbers mentioned above. As was suggested in section 6.1., numbers, and diagrams based on numerical data, are only useful if they convey exact ('true', 'meaningful') information.

Some numerical statements may be untrue due to manipulation errors. Thus, Shackley states:

> The shape of the histogram depends on the class interval used, and Figure 8.3 shows the appearance of two histograms constructed from the same set of data [classed particle size distribution] but using different class widths. (1975:94)

However, the figure referenced shows two markedly disparate distributions, despite the reiterated claim in the caption of their equality. More complex non-information may be produced by

reliance on an inaccurate data transformation procedure. Shackley gives the tabulated results of a Coulter Counter particle size analysis of two samples of Hamwih (Southampton) Brickearths (ibid., Fig. 12.2, p.143). Apart from the fact that equivalent particle diameter is said to be in metres when it is in microns, the table is clear. These results are then plotted on arithmetic probability paper as cumulative weight percentage curves (ibid., Fig. 12.3, p.145). However, these curves do not accurately represent the data tabulated in Shackley's Fig. 12.2. Shackley then states:

> The Inclusive Graphic Statistics of Folk and Ward (1957) were then calculated from percentile values read from the curve, and are summarised below in Table 12.1. (ibid., p.144)

The present author calculated a mean (Mz) for one of the samples ("sample A"), of c.5.33Φ (c.25 microns) from the curve of Shackley's Fig. 12.3, and of c.5.70Φ (c.19 microns) from the accurately replotted data of Shackley's Fig. 12.2; the value given by Shackley in her Table 12.1 is 0.40Φ (c.758 microns, that is, Shackley is suggesting that a 'brickearth', akin to a loess, has a mean particle size in the coarse sands. Similarly, the kurtosis of this sample is around unity (as in a normal distribution) whilst Shackley gives a figure of 6.557, which is almost unknown in natural sediments! It is not at all clear from this information exactly what Shackley is doing wrong. However, if we turn to the particle size analysis of the Newport (Isle of Wight) Sand in a section entitled "Automated Data Processing" (ibid., pp.101-105), some light is thrown on this problem. Raw data from a sieving analysis was fed into a computer program called "SIEVETTE"; this program produced the printout shown in Shackley's Fig. 8.9. The program claims to have derived the percentile values needed for the Inclusive Graphic Statistics by "linear extrapolation". The present author presumes that what is meant is 'linear interpolation,

using a log-normal scale on the abscissa and a gaussian transformation of percentages on the ordinate'. However, the program gives the 95 percentile, for instance, as 3.757Φ, whilst a more accurate interpolation from the raw data gives a value of 3.65Φ. The program goes on to derive statistics based on the inaccurate percentiles. The skewness is given as 0.000, whilst even using these inaccurate percentiles a value of -0.172 is provided by the Folk and Ward formula. We may conclude that the program "SIEVETTE" is incapable of doing simple arithmetic. The situation is not improved by the fact that the graphic representation of this data, Shackley's Fig. 8.8, which was presumably drawn by hand, again fails to represent either the original raw data or the incorrect data produced by the computer program.

The simple lesson to be learnt from the above examples is that one should check, and recheck, one's own data, just as one should do for anybody else's data. The more complex a transformation of primary data becomes, the more scope there is for error, and the less likely it is that the final value will be expressed on a scale that will easily allow intuitive recognition of an error. Shackley, the author of the standard texbook on sediments for archaeologists, was obviously not sufficiently experienced with particle size analyses to notice that the apparent kurtosis of the Hamwih Brickearth (coarse sand?) was practically off the scale, since she went on to make the inference:

> Such high kurtosis values could apply to either estuarine or aeolian sediments .... (ibid., p.144)

Transformations, in the form of indices, are commonly used in cave sedimentology. For example, the concept of rounding of limestone clasts has been quantified in various different ways.

Laville (1973, 1975) uses a method, described by Bonifay (1956), which entails subjectively dividing a sample of clasts into four rounding categories, calculating the weight percentage in each class, multiplying the results by a coefficient and then adding the four values to give an index. First, the question of which sample of clasts to take must be decided. Laville uses the best represented class(es) of 'stones' (that is, somewhere in the range 1-10cm), a factor which presumably will vary from layer to layer. Since size of clasts might be expected to interact with rounding, it would probably be better to reduce the effect by using percentages based on number counts rather than on weight. Then, the class limits must be set. Laville selects four clasts from a whole site which are typical of four 'fields' on the continuum angular-rounded, and which then act as reference material; each clast in a sample is classed depending upon its resemblance to one of the reference pieces. This index is therefore site-specific and one cannot compare values between sites. The coefficients used are 0 for unrounded clasts, $\frac{1}{3}$ for slightly rounded clasts, $\frac{1}{2}$ for rounded clasts and 1 for very rounded clasts. The index therefore has the set limits 0 (totally angular) and 100 (totally rounded). However, the choice of the coefficient $\frac{1}{2}$ for the rounded class (instead of the more logical $\frac{2}{3}$) means that the scale is not symmetrical (e.g. if the four classes each contain 25% of the sample, the index value is not 50 but 45.83). Finally, the index gives no idea of spread or skewness (although it does in fact react in a bizarre manner to these parameters because of the illogical coefficient $\frac{1}{2}$; e.g. the figures 50%, 0%, 0%, 50% give an index value of 50, whilst the figures 0%, 50%, 50%, 0% give a value of 41.67).

Miskovsky (1972), referring to 'corrosion' (defined as

rounding of angles), uses a similar index to that employed by Laville, but the scale has been made symmetrical by the use of equal interval coefficients. Five classes are defined, with the coefficients of 0, 1, 2, 3, and 4 and the result divided by 100, to give an index in the rather peculiar range 0 (totally angular) to 4 (totally rounded). Miskovsky calculates this index on two standard size classes, 1-2cm and 2-3cm. He also makes a rather obscure statement which leaves us wondering what sort of percentages are used:

> Nous avons créé un indice de corrosion en affectant d'un coefficient différent le nombre et le poids de chacune des catégories .... (1972:113)
>
> [Trans.: We have created an index of corrosion [rounding] by assigning a different coefficient to the number and the weight of each class ....]

As well as the index, Miskovsky actually plots the five untransformed percentages of each class for clasts in the range 1-10cm, but again it is not clear whether number or weight percentages are being used.

Brunnacker and Streit (1965) use a four class description of edge-rounding of dolomitic clasts. Rather than splitting a sample into rounding classes, they first split it into four size fractions (1-2, 2-5, 5-20 and 20-50mm). Then each size fraction is described as being of one dominant rounding class. An empirical pattern emerges from their data, such that the finer grades are nearly always more rounded than the coarser grades. An appreciation of rounding is gained from the diagrammatic results by observing how 'quickly' the normal, 'mostly rounded' state of the finest clasts tends towards sharper edges on coarser clasts. No numerical index is calculated, but a diagram with the weight percentages of each size class, for each sample, is displayed alongside the rounding data. Such a presentation is simple and easy to understand, but it is dependent upon the clear size patterning of

rounding. If one size class showed great variation in rounding, the fact could not be represented by this method.

Butzer (1971:167) suggests the use of the Lüttig (1956) index of rounding, which is the percentage of smooth, convex circumference of a clast (or usually of its most stable projection). This method is only really useful in the case of quite extreme variation in rounding (e.g. comparison of slope scree and water-worn material). Butzer's advice concerning the presentation of data is interesting. First, he calculates a mean value of the index over all clasts in a sample. Depending upon where the mean falls, the sample is allocated a verbal description on a scale of five (0-10%, angular; 11-20%, subangular; 21-40%, subrounded; 41-60%, rounded; >60%, well-rounded). No mention is made of the size class of clasts to be used. Butzer goes on to calculate the standard deviation (s) of the Lüttig index values, which, weighted by the mean ($\overline{x}$), gives a percentage coefficient of variation:

$$CV\% = 100s / \overline{x}$$

The CV% values receive verbal descriptions (0-25%, very homogeneous; 25-50%, homogeneous; 50-75%, heterogeneous; >75%, very heterogeneous). Butzer says nothing more about the behaviour of either the mean or the CV.

Since the class intervals of the verbal descriptions of average rounding are not equal, we may conclude that the scale of the Lüttig index is probably geometric, with much finer definition being achieved nearer the angular end of the scale than nearer the well-rounded end. Butzer's use of the CV to express 'homogeneity' stems from a misunderstanding of this parameter, aggravated by a superficial resemblance between its behaviour under these

circumstances and the behaviour to be expected from a measure of 'spread' on a geometric scale. There is a universal tendency in numerical data on unrestricted scales for the standard deviation to be directly proportional to the mean (e.g. a group of wide-trunked trees will have a greater standard deviation than a group of thin-trunked trees, etc.), without there being any algebraic dependence between the two parameters. The CV is therefore a convenient way of stating standard deviation relative to the mean, given an unrestricted scale. It is obvious that the CV is only valid if the standard deviation is independent of the mean. The Lüttig index is a percentage, so that the standard deviation is automatically (artificially) dependent upon the mean. Both the standard deviation and the CV are increasingly warped as the mean approaches either limit of the scale. Butzer's division of the CV 'percentage' range into what appear to be four equal-interval classes suggests that a true measure of 'spread', on an arithmetic scale, has been achieved. Note, however, that the "very heterogeneous" class is defined as ">75%", not as "75-100%". This is because the upper limit of the CV parameter, as defined by Butzer, is not 100, but infinity. The scale of the CV, calculated on percentage data, is not arithmetic but exponential, with mathematically (although not logically) valid classes requiring separate definition for every possible mean! All these problems stem from the fact that the standard deviation of percentage values (the original Lüttig index values) is an extremely obscure parameter. Indeed, so is the mean in this case. A much more reliable measure of spread is achieved by transforming the Lüttig scale into a true geometric scale and calculating some such statistic as the interquartile range. If a value of the spread statistic is in error due to the finite limits of the

scale, this will immediately show in a graphic plot (e.g. a frequency histogram) as marked skewness of the whole distribution. Methods of correcting the effects of finite scales will be found in most good statistics textbooks. If there is marked skewness in a plot of transformed Lüttig index values, even when few values fall near the limits and there is no obvious sign that the sample represents a mixing of sources, then the optimal geometric transformation has not been used. Intuitively, it is probable that different rock types will have different optimal transformations.

Farrand (1975) has provided a well balanced discussion of rounding of limestone clasts. First, he dismisses many standard procedures (implying such indices as that of Wentworth (1919), involving measurement of the curvature of the most angular corner) as not being applicable to immature limestone clasts, since such clasts have many corners and considerable variation in angularity. Farrand uses four roundness classes, but wisely declines to calculate an index on the grounds that all idea of the spread of values would be obscured by so doing. The difference between Farrand's approach and that of other researchers is that he attempts to give a more formal (although still necessarily subjective) definition of the four classes: Class 1, only angular corners; Class 2, angular corners more common than rounded corners; Class 3, rounded corners more common than angular corners; Class 4, only rounded corners. Moreover, he has tested this method for reproducibility, with both the same operator and different operators, and has found the values to be satisfactorily stable. It is a pity that Farrand has spoilt this simple and logical approach when it comes to presenting the data. In the text (1975:36), he says that stones in each size class were

evaluated for rounding, but he does not say which class or classes were used to give the results shown in the single diagram (ibid., Plate 1). Nor does he say whether number or weight percentages were used.

Shackley (1975) figures and uses (with flint and chert samples) the Powers (1958) visual comparator chart for rounding. This chart is based upon a two-dimensional approximation of the Wadell (1933) rounding index: the sum of the radii of all corners in a projection, divided by the product of the number of corners and the radius of the maximum inscribed circle. The chart has nine classes, defined on a geometric scale of Wadell index values, since it has long been noticed that slight differences in rounding are much easier to assess on angular particles than on rounded particles (cf. Pettijohn 1975). Since rounding is theoretically independent of shape, each rounding class in a chart of this type should illustrate various different shapes to help the eye to ignore the shape factor. The Powers chart is perhaps not quite 'universal' enough with respect to the shapes figured. A researcher working with a particular type of limestone (or other rock type) could produce his own comparator chart with more relevant shapes. Griffiths (1967) has indicated that such charts are open to significant operator error (subjectivity), but the present author is not convinced that direct measurements of curvature and the like are any more accurate (cf. p.523).

Rounding was chosen to illustrate the complexities of apparently simple summary statistics precisely because it is a difficult property to define. The examples given above are only a few of the ways in which this problem has been approached. It should be clear that the reliability of such indices depends upon a logical argument, starting from the property and arriving at

the numerical index. Far too often, indices have been defined without proper consideration of how values may be affected by the algebra rather than the property of interest.

Indices of rounding provide values which may be considered either as comparative (between samples) or absolute, but which do not automatically imply a specific sedimentological correlate. Many indices are specifically designed, not primarily as mere summary statistics, but as assumed indicators of sedimentary processes. These indices derive their significance from repeated observation. Thus, Butzer (1971) gives values of the indices width/length and width/thickness of pebbles which are commonly observed associated with, on the one hand, a rolling motion, and on the other, a sliding motion during transport. Similarly, Passega (1964) has shown that 'C/M' diagrams (grain size of the coarsest percentile plotted against median grain size) are useful in recognising mode of transport. Both these examples employ comparatively simple arithmetic devices that do not have an undesirable effect upon the values of the indices. Furthermore, in both cases, there are good theoretical (physical) reasons for expecting the indices to behave in the way actually observed. However, as was stated above, an index is defined for a specific purpose and has no meaning beyond that purpose. Thus, the width/length index of clasts in general does not have the same significance as the width/length index of 'pebbles', since, in this context, we may define a 'pebble' as a clast which, through attrition or sorting, has demonstrably reached shape equilibrium with a transport process. The width/length ratio of clasts that have not undergone transport may be of interest for some other reason, but it obviously tells us nothing about transport processes. It is very easy to accept the fallacy that, because a calculation

can be performed, the significance of the calculation, as recognised in a specific context, becomes universal. In the terms of our example, the fact that the width/length ratio of a clast can indeed be measured clearly does not prove that it has been transported. Nevertheless, in other areas (especially with respect to particle size analysis), such logical errors are very easily made and often difficult to recognise once made.

Particle size analysis has already been mentioned in the context of sampling (section 6.2.). Here we are concerned with scales of measurement and summary statistics. Most particle size analyses start from a graphic plot of the raw measurement data. Particle diameter is plotted on the abscissa (increasing to the left in most English language texts, but often to the right in German and French language texts). Frequency, usually as a weight percentage, is plotted on the ordinate, with zero at the bottom. These axes should not be rotated (contra Shackley 1975, Fig. 8.7), since the independent variable is conventionally plotted on the abscissa of any graph. This convention is important in this context because an operator quickly acquires an intuitive appreciation of the 'meaning' of particle size curves, an appreciation which is confused if scales are inverted.

It has long been recognised that, if particle diameter is plotted directly against weight percentage, a violently skewed distribution results, no matter what natural sediment is used. This is a general property of sediments, predictable from the observation that volume is a power function of diameter. Such plots are very difficult to work with and some method of removing as much of the skewness as possible is called for. This may be done by converting the particle diameter from an arithmetic to a geometric scale. The choice of a geometric scale was

originally based upon empirical observation and ever since the work of Udden (1898) a base-2 scale has been used. Note that this is merely a graphic convention which allows appreciation of finer differences in distributions; no intrinsic 'meaning' is implied, although there is, in fact, a rationale, of no importance here, for this particular transformation in geometrical terms. There are two basic systems currently in use. The French system involves the use of logarithmic graph paper; results are usually published without the grid but with a series of conventional points (i.e. 2, 1, 0.5, 0.2, 0.1, 0.05, 0.02, etc.) in millimetres marked off at geometrically related intervals along the abscissa. It is therefore very difficult to interpolate diameters, since here the scale is explicitly logarithmic. The more common system, which hopefully will soon become universal, involves the transformation of diameters to logarithmic values which can then be plotted on a simple arithmetic scale. The new scale is not in millimetres, but in Φ (phi) units (Krumbein 1936):

$$\Phi = -(\log_2 mm)$$

Thus, 2mm = -1Φ, 1mm = 0Φ, 0.5mm = +1Φ, 0.25mm = +2Φ, etc. The sign convention stems from the fact that most sediments will have a greater spread below 1mm than above it, so that it is more convenient to invert the sign of the logarithm (contra Shackley 1975:91, who states that the logarithm should be multiplied by "-0"!). Because base-2 logarithms are not readily available, it may be easier to work with base-10 or natural logarithms:

$$\Phi = -(\log_{10} mm / 0.30103) = -(\ln mm / 0.6931472)$$

Published conversion tables between Φ and mm are often significantly inaccurate (cf. Shackley 1975, Table 8.3), especially for the finer grades, and are best avoided. It is important to understand that the French system and the Φ system

are algebraically identical, that is, a distribution plotted in one system will have an identical shape when plotted in the other. The greater utility of the Φ system lies in the fact that interpolation, which will be necessary for the derivation of summary statistics, is conceptually more simple on an arithmetic scale (although it is not intrinsically more accurate).

The weight percentage scale on the ordinate may appear in three different forms. When an arithmetic scale is used, it may be a simple frequency, resulting in a roughly bell-shaped plot, or a cumulative frequency, resulting in a roughly sigmoid plot. If, however, cumulative frequency is plotted on a gaussian ('probability') scale (cf. Hatch & Choate 1929), there is a tendency for the distribution curve to become more like a straight line (infra). Use of a gaussian scale of percentages again makes it very difficult to interpolate values, even when a basic grid of modules is figured. Although the present author has never seen it done in print, it would be much easier to convert the transformed percentage values back to an arithmetic scale:

$$\frac{p}{100} = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz$$

where p = percentages

z = standard normal deviate

(the unit of the new arithmetic scale)

This integration may be evaluated by various approximate methods (e.g. the method of rectangles) but the values of the cumulative normal frequency distribution tabulated in most statistics textbooks (to at least two decimal places) will usually suffice. The z-values corresponding to the percentiles used in most graphic summary statistics (infra) are as follows: 5% = -1.64z; 16% = -0.99z; 25% = -0.67z; 50% = 0z; 75% = +0.67z; 84% = +0.99z;

95% = +1.64z. The z-scale is symmetrical about zero (50%); it has no finite limits but it is not usually necessary (indeed, it will represent spurious accuracy) to plot values outside the range -3 to +3 (0.13% to 99.87%). Using the transformed variables Φ (for diameter) and z (for cumulative weight percentage), there is no need for special graph paper to plot particle size distributions and, once the operator has become familiar with these variables, the scope for plotting and interpolation error is considerably reduced (cf. fig.16).

Once a particle size distribution has been plotted, various summary statistics may be calculated. Indeed, some statistics may be calculated without a graphic plot, but the present author recommends that a plot always be made, both because this provides a check against numerical errors and because summary statistics do not, by very definition, fully characterise a distribution. There are four commonly used statistics: mean, standard deviation (spread, sorting), skewness (asymmetry) and kurtosis (peakedness - low values indicate 'flat-topped' distributions with only minor 'tails', and high values indicate 'peaked' distributions with rather long 'tails'). There are many different algebraic definitions of these statistics (cf. Griffiths 1967), but the most commonly used in English language texts are those of Folk and Ward (1957). Such statistics are today calculated with arithmetic cumulative percentages and logarithmic <u>units</u> (as distinct from scales) of diameter. Many French authors neglect to point this out, leaving the reader extremely puzzled. For instance, Miskovsky (1972) defines a simple index of interquartile sorting (<u>hétérométrie</u>):

$$Hq = (Q_3 - Q_1) / 2$$

The third quartile is the diameter at 75% and the first

that at 25%. Since Miskovsky only uses this index for material under 2mm, the index Hq can never be greater than unity if one simply reads off the particle diameters (in mm) on the abscissa of his graphs. However, all but one of the values of this index shown in Miskovsky's Fig. 22 are greater than unity. Particle diameters must be converted to Φ units before the calculation of this index, although Φ units are never mentioned, either explicitly or implicitly, by Miskovsky at any point in his text.

Consider a theoretical distribution of particle sizes which is log-normal. This means that, when particle diameter (on a logarithmic scale such as Φ) is plotted against frequency, a perfect bell-curve will result. The mean will be at the mode (the highest point of the curve); the spread is a variable, dependent upon the precise distribution, but independent of the mean; skewness will be zero (the curve is perfectly symmetrical about the mean); and kurtosis will have a value characteristic of a normal curve (dependent upon the algebraic definition used, e.g. 1.0 for the Folk and Ward statistic). If the distribution is then plotted as a cumulative arithmetic frequency, a perfect sigmoid curve will result, such that one half of the curve will be the image of the other half, rotated by 180° in the plane of the paper. If the distribution is further transformed to a cumulative frequency on a gaussian, or z, scale, a perfectly straight line will result. The values of the summary statistics do not, of course, change during these transformations.

It is often suggested that real particle size distributions may be approximately log-normal. This is a useful generalisation but we will need to examine this proposition more carefully here. A reason for the logarithmic scale of sizes has already been given above (the power relationship between volume and diameter).

But why should a distribution be normal? Normal distributions are often associated with what may be called 'goal-seeking' behaviour. A system is 'trying' to achieve some ideal state; most often it is quite 'successful', whilst more rarely it is much less 'successful'. The sum of all the 'tries' is a normal distribution and the system is said to be in statistical equilibrium. The mean may be thought of as the 'ideal state' and the standard deviation represents the overall 'success rate'. This is a result of the basic theorems of probability, since the normal distribution is merely the continuous form of the binomial distribution, itself the first 'derivative' of probability theory (for further discussion, cf. Mood & Graybill 1963; Feller 1968). We might therefore expect a log-normal particle size distribution to be the result of a discrete process in statistical equilibrium. Since distributions are never perfectly log-normal and processes are never perfectly discrete, we may suggest that the more closely a distribution approximates to log-normality, the nearer we are to observing a discrete process.

In reality, very few sediments are even approximately log-normal, although some dune sands come quite close. This statement appears to contradict what is said in most textbooks. It is really a matter of the degree of approximation that is acceptable <u>for a given purpose</u>. What is certainly true, however, is that many sediments are made up of mixtures of a few sedimentary populations, each of which may be a very good approximation (by anybody's standards) to a log-normal distribution. This was first realised by Moss (1962, 1963), who noted that most water-laid sediments are made up of three main log-normal populations, referable to the processes of rolling or traction across the bed, saltation, and transport in suspension. Theoretical considerations

of the physics of transport and deposition produce very similar results to those inferred from the observation of real sediments, and not only in subaqueous contexts. It would therefore appear that a process such as saltation is quite close to an 'ideal' discrete process. Indeed, a dune sand which has a total distribution close to log-normality will be made up almost entirely of a saltation population. The analysis of log-normal populations in a single sediment is now a common approach (cf. Visher 1969), but there are some restrictions which are of major importance when it comes to the types of sediments usually found in caves, and in many other kinds of archaeological site. A log-normal population of particle sizes represents an equilibrium with a process. However, when a new process starts to operate or when exotic sediment comes within the reach of an existing process, it takes time for a new equilibrium to be achieved. The time factor depends largely upon the relationship between the energy of the process and the resistance to movement of the particles (weight, shape, cohesion, etc.). Thus, a 'weak' process, acting in an environment where a wide range of particle sizes are becoming available comparatively regularly, will achieve equilibrium only for short periods, over a restricted particle size range and in very small areas (perhaps only a few centimetres across). The chances of finding log-normally distributed material under such circumstances are minimal.

A sediment that is made up of several log-normal populations (a 'composite' log-normal distribution) will not show an overall log-normal distribution. Only the mean and standard deviation are required for a log-normal distribution to be completely and accurately described. For a composite log-normal distribution, skewness and kurtosis measures are also required but,

even with these additional statistics, the distribution is not completely described. Note that none of the four statistics has any explicit relationship with any of the component log-normal populations; the statistics refer to the mixture, not to the parts. Composite log-normal distributions often deviate from overall log-normality in specific ways. For instance, beach sands are often negatively skewed, with a prominent 'tail' of coarser grains (due to an important traction population and progressive winnowing away of the fine suspension population). Knowing the pattern of a total distribution, defined by the four main summary statistics, may sometimes permit similar conclusions to those achieved by the study of component populations. Indeed, this was the main approach up until the late 60s (cf. Friedman 1961), and it is still useful under many circumstances. However, the approximate nature of the summary statistics is underlined by Glaister and Nelson (1974), who conclude that a study of the component log-normal populations is more likely to lead to significant observations that are related to processes and environments.

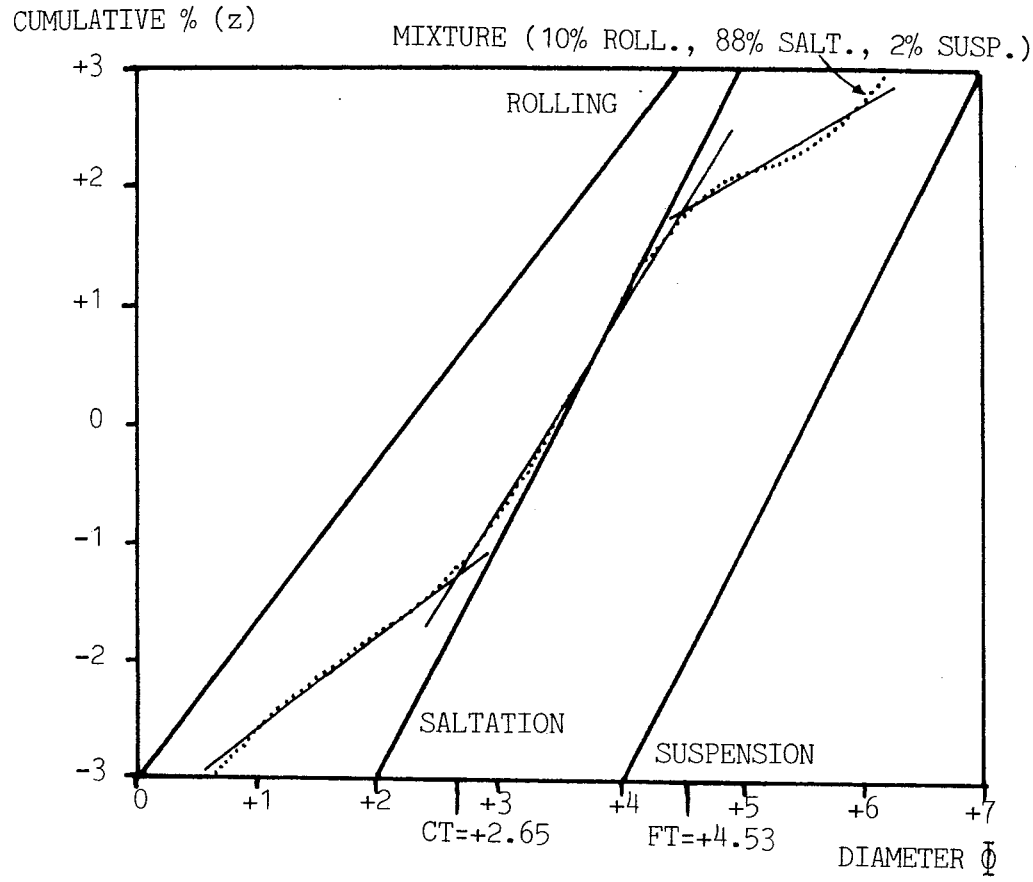
Before continuing, it should be noted that log-normal populations, although probably the most common of the equilibrium forms, are not the only populations likely to occur. For instance, populations approximating to a Rosin distribution (Rosin & Rammler 1934; the cumulative frequency scale looks something like a gaussian scale, but with the first 50% 'compressed' and the second 50% 'distended') are known in materials that have not been transported but which are the result of simple breakage of larger particles into smaller ones. This distribution might be applicable to limestone screes, although, to the present author's knowledge, the careful sampling that would be necessary to prove the point has never been carried out, despite the fact that this distribution

has often been found in artificially crushed limestone.

The question now arises of how to recognise log-normal populations. It was stated above that a log-normal distribution plots as a straight line on a graph of Φ units against gaussian cumulative frequency. However, no other distribution, even a composite log-normal distribution, has this form.

> A plot of grain size-curve on log-probability paper gives straight-line curve segments, each segment denoting a single subpopulation ['population' in the terms of the present text]. Each subpopulation shows a log-normal distribution and differs from other subpopulations in median and standard deviation. The number, amount, size-range, mixing [overlap], and sorting of subpopulations of a sample vary systematically in response to provenance, sedimentary process, and sedimentary dynamics .... (Reineck & Singh 1980:137)

The above statement is highly misleading and it is worthwhile proving the point, especially since similar statements will be found in other textbooks that, like the one by Reineck and Singh, are otherwise extremely reliable. Fig.16 represents a theoretical particle size distribution made up of three perfectly log-normal populations. The exact parameters are arbitrary, chosen for ease of graphic illustration, but the 'sediment' does in fact resemble sandy material often found in marine shoal environments. The three straight lines in the upper graph show, independently, the three log-normal populations (labelled 'rolling', 'saltation' and 'suspension' to continue the marine analogy). The mean of each population (read simply at Oz) is plausible for the processes suggested. The exact values of sorting are not relevant to the argument, but they are proportional to the slope of each line. Thus, the 'saltation' and 'suspension' populations have the same standard deviation, whilst the 'rolling' population is less well sorted. The individual points plotted on this graph represent the overall sediment, resulting from a blend of the three populations in set proportions: 10% 'rolling', 88% 'saltation' and 2% 'suspension'. Even in this 'sediment', which has a



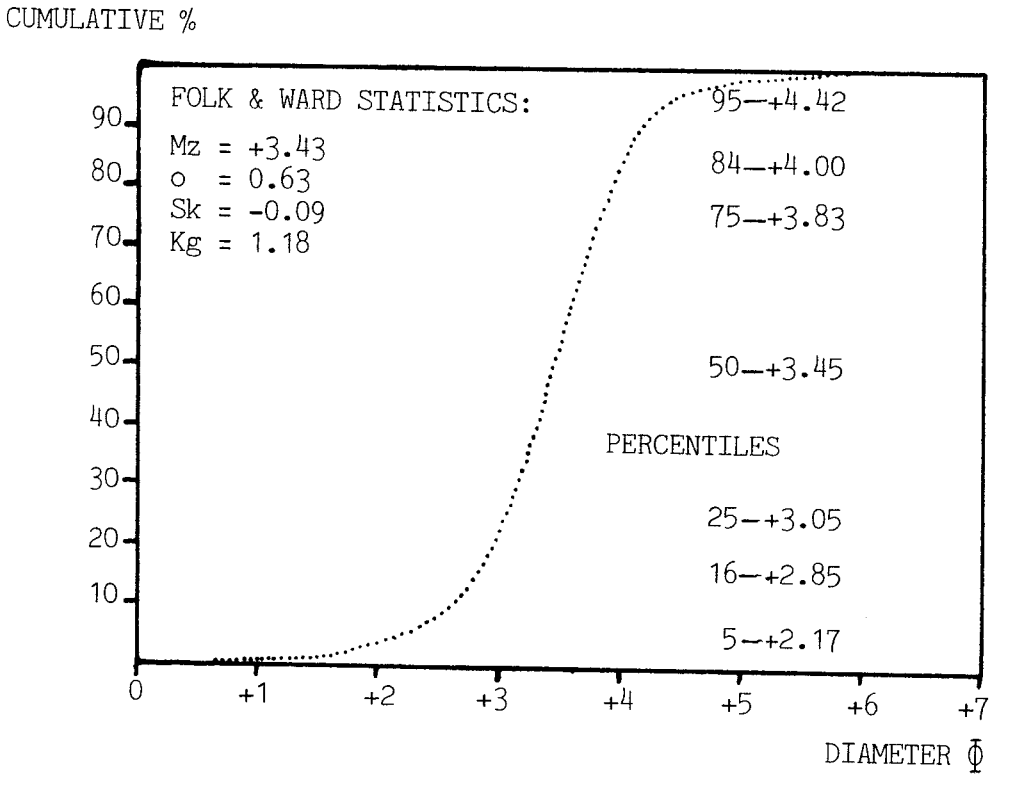


Fig.16 Theoretical Particle Size Distribution

dominant 'saltation' population, it is clear that the curve is not made up of straight-line sections but of complex curved sections. There are, however, two zones of pronounced inflexion. The usual procedure is to fit the best straight line to each of the three 'less curved' sections (as has been done in this graph). The points at which these straight lines intersect are known as the coarse truncation (CT) and the fine truncation (FT), referring to the way in which the central 'saltation' population is truncated by the flanking populations (CT and FT values may be given for each population if required). Note that the degree of curvature in each of the three sections is dependent upon the total sediment type (the mean and standard deviation of each population and the proportions of each population present); no inference about the degree of approximation to log-normality of the original populations can be made from the graph (the 'suspension' segment is much more 'curved' than the 'saltation' segment, but both populations are none the less perfectly log-normal). Many (most) natural sediments show much greater curvature of population segments than in this example. The truncation points can be used to estimate the proportions of each population. In this case, we find 10.6%, 86.0% and 3.4%, which are not far from the true values, although this is again due to a fortunate choice of original parameters. An estimate of the means gives +1.56Φ, +3.59Φ and +5.52Φ, which are rather poor approximations of the true means of +2.25Φ, +3.50Φ and +5.50Φ. The slope of the best-fit lines is dependent upon the proportions of the blend. Thus, the 'saltation' population estimate of spread is not too bad, that of the 'rolling' population is worse and that of the 'suspension' population is so poor that it appears to be less well sorted than the 'rolling' population. It can therefore be seen that study of graphic

representations of composite log-normal distributions is far from straightforward. The recognition of a straight-line section on a graph such as this is, in fact, more likely to be proof that a discrete log-normal population is not present, a fact which one should bear in mind when analysing complex, low energy deposits such as many cave sediments. Rather than working directly from a graphic plot of the overall sediment, it is possible to construct an iterative computer program which will try to identify the original log-normal populations. However, there are no unique solutions to such a problem, that is, even if the program is told how many populations are likely to be present, there will be several combinations of populations and proportions which will produce identical overall distributions. Add to this the fact that one or more real populations may be more or less immature and not a true log-normal distribution, and compound this difficulty with measurement errors, and it is clear that it will only be useful to undertake such a computer analysis in cases where the operator is pretty sure of what is going on in the first place.

For the purposes of comparison, the same theoretical 'sediment' has also been plotted in the lower graph of fig.16 as a cumulative arithmetic frequency. The resulting sigmoid curve no longer betrays in such an obvious manner the presence of the component log-normal populations. The Folk and Ward statistics help a little. Kurtosis is higher than in a log-normal distribution, suggesting that the main population is near the centre of the range, whilst the slight negative skewness suggests that the flanking population with the higher proportion is on the coarse side of the main population. Even these inferences are dependent upon the assumption that three log-normal populations are in fact present, a point which is by no means obvious from

the sigmoid graph itself or from any of the summary statistics.

This long discussion of particle size distributions was necessary in order to underline the true nature of the parameters used to describe them. Even in situations where sediments are known to represent equilibria with a few basic processes, the task is far from easy. All the methods described above essentially involve the evaluation of indices. Even the degree to which a curve approximates a straight line is an 'index'. In the cave situation, where most sediments will be complex mixes of immature sediments, blind trust in these methods may be disastrous. One must always keep in mind that an index, when divorced from its proper context, is meaningless.

Tankard and Schweitzer (1976) have used the Folk and Ward statistics to describe the sediments of the cave of Die Kelders (Cape Province, South Africa). They note that the lowest layer (17) has high negative skewness and a comparatively coarse mean; citing Friedman (1961), they suggest a beach sediment. What is important is that this suggestion is backed up by the presence of littoral quartz grain surface textures and of echinoid spines, not to mention a plausible argument as to why the sea should have been close at that time.

Shackley (1978) gives the same four summary statistics for the sediments of Devil's Lair (Western Australia), a limestone cave that now lies some 5km from the sea. There is no reason, either inferential or based on clear marine indicators, to suppose that the sea was close to the cave during the deposition of the sediments or for a considerable period before their deposition. Shackley mentions no ancient marine deposits in the area. Nevertheless, she states:

> However, the existence of 4 samples [of non-carbonate residues] with

> negative skewness values is interesting, since this feature has been taken by many workers (for example Friedman 1961) as typical of beach sands (1978:36)

Neither Friedman nor anyone else, to the present author's knowledge, has ever claimed that negative skewness is indicative of beach sands. Indeed, Friedman reports dune sands and fluvial sands with negative skewness comparable to Shackley's samples. One of Shackley's samples is a "poorly sorted, sandy gravel" and has statistics (Mz 0.15, σ 1.67, Sk -0.12, Kg 0.68) that show it to be completely outside the range of sediments studied by Friedman. These deposits obviously represent mixed and/or immature sedimentary stocks, and the conjecture as to a marine origin is a red herring, resulting from facile treatment of complex indices.

Returning briefly to the discussion of Die Kelders by Tankard and Schweitzer, many of the layers above the marine deposit are predominantly composed of aeolian material. This is deduced from the context of the site and also from the rounding, polish and general size distribution of the particles in the sediments. However, some layers contain minor components of coarser shell and finer "mud-humus", components that affect the overall summary statistics. The discussion of these sediments by Tankard and Schweitzer is relatively clear and their conclusions are convincing. Nevertheless, since they state that all size distributions were initially plotted as cumulative frequencies on a gaussian scale, it is a pity that they did not figure some of these plots because the different populations would probably have shown up rather well, thus avoiding the need for much of the long explanation of the behaviour of the summary statistics.

At Die Kelders, the sediments are basically representative of comparatively simple, high energy environments. In low energy cave deposits, the Folk and Ward summary statistics may help to

describe the particle size distribution, but the sedimentological significance is often far from obvious (cf. Collcutt 1975; T.J. Lawson 1982).

Rivière (1952, 1957) has developed a system for describing grain size distributions that is extremely obscure, although, as far as the present author is able to tell, mathematically accurate. Rather than converting a size distribution into a convenient form through standard scalar transformations, Rivière starts with a cumulative frequency on an arithmetic scale (a sigmoid curve), which may be described by a general formula:

$$y = \int ax^n \, dx$$

where y = cumulative frequency

a = constant

x = dimensionless number representing particle diameter

n = an exponent.

The values of x are found by:

$$x = d / dm$$

where d = particle diameter

dm = smallest particle in range.

The entire range of a sediment, G, is therefore:

$$1 <= G <= dM / dm$$

where dM = largest particle in range.

By transforming a distribution in this way, Rivière produces a standardised curve for any distribution. Note that, unlike the case of a normal transformation, this technique has no a priori link with any suspected distribution property of a real sediment. It is merely a mathematical transformation designed to produce two indices: g = logG, and n. The index g is called the 'index of sorting', and is merely a simple logarithmic

transformation of particle size range. The index n has no explicit sedimentological meaning. However, the empirical observation of certain types of sediment led Rivière to call n the 'index of evolution (maturity)', since it varies proportionally to this subjective property. The actual application of this technique is complex, involving the use of canonical curves (or a computer) to estimate parameters; this is not important to the present discussion and any interested reader is referred to the works cited above. What is important, however, is the type of sediment for which these indices were designed. First, any sediment analysed must have a negligible probabilistic character, that is, one must be sure that any sample is an almost perfect representative of the true overall population. This effectively restricts the technique to fine sediments. Second, any sediment analysed must be unmixed:

> Il résulte de cette étude (qui exclut le cas de mélanges brusques de stocks d'origines et d'évolution différentes) .... (Rivière 1952:166)
>
> [Trans.: It follows from this study (which excludes cases of sudden mixing of stocks from different sources and with different degrees of evolution [maturity]) ....]

These indices are again designed for environments with one, constant set of processes and one, constant set of sedimentary sources. The indices are even more misleading than the Folk and Ward summary statistics when applied to mixed or very immature sediments. This can be proven mathematically. In a valid sediment, the index n can be calculated on any <u>part</u> (defined by choosing dm and dM arbitrarily within the total range of particle sizes, and redefining this part as 100%) of the sediment distribution and it will remain constant. In an invalid sediment, the index n will vary according to which part of the sediment distribution is analysed. This follows simply from the requirement that a valid sediment distribution, with a logarithmic particle

diameter scale, will be exactly defined by the general integration given above (cf. Rivière 1957). This purely mathematical artefact appears, to the present author, to coincide approximately with the fact that some sediments contain log-normal populations, although Rivière bases his arguments on empirical results and never makes this underlying link explicit.

Because of the obscurity of the mathematics, the present author cannot see that Rivière's indices can ever have any advantage over the established methods of description. Indeed, no clear definition of 'maturity' is ever given. In cave sediments, however, the use of the index n will nearly always be totally misleading. At least negative skewness, for example, actually means something in terms of a real distribution, whether or not we go on to infer a marine environment or whatever. The index n has no explicit meaning as a purely summary statistic (it does not describe the shape of a curve in easily comprehensible terms) and the empirical concept of 'maturity' cannot be applied when a distribution fails to comply with the mathematical restrictions imposed by the method of calculation. However, precisely because the index n is so obscure, it is difficult to recognise that it is applicable only under very special circumstances. Thus, Miskovsky (1972, 1974) has used the index, with its implication of 'degree of maturity', for all sorts of cave sediments, apparently in total ignorance of its true nature.

Incidentally, the only case known to the present author of the use of a gaussian transformation of cumulative frequency to study blended quasi-log-normal populations in the sediments of a European archaeological context is furnished by H. Kadiri and M. Leger (in Taborin et al. 1979), working at the open-air site of Etiolles (Essonne). These are mature fluvial sediments and

the technique thus provides valid and informative results. It is to be hoped that the sedimentologists' efforts will not be wasted and that the archaeologists will use this unusually sensitive data to help interpret site formation processes.

One last point must be made concerning particle size distributions and the effects of setting arbitrary size limits. Most researchers do this and it may often be strictly necessary in practice. For instance, Miskovsky (1974) gives curves for the sand fraction alone and Laville (1975) gives curves for the fraction under 2mm. Summary statistics may be calculated from these curves, but it must never be forgotten (even with 'high energy' sediments) that the statistics apply only to the material within the limits. As an example, if a theoretical log-normal size distribution is plotted so that the great majority of the distribution falls within two arbitrary limits, summary statistics for the curve will show a true mean and standard deviation, together with zero skewness and a (statistically) normal value of kurtosis. If the range between the limits is kept constant but the limits are moved to higher or lower values of particle size, so that increasingly greater proportions of the distribution fall outside one of the limits, increasingly inaccurate values of apparent mean and standard deviation, together with increasingly non-normal values of skewness and kurtosis, will result. This limit effect applies to any distribution, not just to log-normal ones. Before drawing any sedimentological conclusions from summary statistics, it is important to consider whether any limit effect is likely to have influenced the statistics and whether this is of importance in the context of particular problems. Consider, for instance, a loess, which will probably have a slightly positive skewness and poor sorting, although the sand

component of the same loess will be negatively skewed and well sorted, perhaps resembling a fine beach sediment.

It is not necessary to discuss the application of statistical tests to cave sediment data (for general coverage of geological statistics, cf. Miller & Kahn 1962; Koch & Link 1970, 1971; Davis 1973; Till 1974 [with reservations, since this text contains many 'hidden' typographical errors]). Such tests are straightforward and depend for their validity only upon the quality of the reasoning which precedes their application. The core of all statistical work consists of relevant hypotheses, a relevant sampling plan and a relevant measurement technique, subjects discussed in this and other chapters of the present text. Converting logic into algebra is only a 'trick of the trade', which should impress but not frighten non-statisticians. If a logical train of argument, between the geological reality and the numerical data, indices or test results, cannot be identified by the reader with relative ease, it is usually because no such logical train of argument actually exists.

7. ANALYTICAL TECHNIQUES

7.1. Laboratory Analysis

It is not the purpose here to undertake even a general discussion of the enormous number of laboratory techniques now available to the sedimentologist. Details of such techniques may be found in many textbooks and papers (cf. Cornwall 1958; Milner 1962; Müller 1967; Carver 1971; Hesse 1971; Shackley 1975; Catt & Weir 1976; Bowles 1979).

The attitude taken here is that 'the end justifies the means'; once it has been decided what information is required to solve a problem or prove a point, any available technique may be used. The only laboratory technique that the present author uses for each and every sample, as a matter of course, is scanning under a binocular microscope. This may be considered as merely an extension of the observations already carried out on site, although it may of course develop into a much more specific analysis. Thus, laboratory analysis is here seen as a problem solving tool.

Compare this with the approach used by such researchers as Laville (1975). A fixed battery of techniques is applied systematically to all samples. Perhaps ninety percent of the researcher's commitment to a site is concentrated in this expenditure of time and effort. Thus, laboratory analysis is seen, in this case, primarily as a data collection procedure.

The present author feels that his approach is more appropriate, given the precise situation in which his research was carried out. However, the second approach should not be discounted out of hand because it has one major advantage over the opportunistic procedures adopted here, namely its very standardisation. Because 'mass' laboratory analysis produces a strictly comparable data set for each different deposit in each different site, it should eventually become possible to explore regional and temporal trends in the basic sediment parameters that would not otherwise be apparent. Furthermore, such data permit a detailed assessment of the relevance of the analytical techniques themselves. Ideally, a combination of the two approaches, with selected topics to be studied by systematic laboratory analysis, would be the best arrangement for the future. Too much laboratory work leads, however, to massive redundancy; it is not necessary to carry out a particle size analysis merely to prove that a sediment indeed represents a stream deposit, for instance.

The question of the general relevance of laboratory analysis, its relationship to fieldwork and the manipulation of numerical data derived from it, has already been discussed in preceding chapters. Here, we are concerned with the actual 'mechanics' of a given analysis. Any analysis involves cost, time, equipment and expertise, all of which are at a premium. In addition to this purely materialistic fact, archaeologists have achieved a certain measure of confidence, based upon experience gained during excavations. Not unreasonably, it is often argued that 'specialists' in the geological sciences, even when they can find the time to help, lack insight into the archaeological problems which are of ultimate interest. Thus, cave sedimentology has grown up in archaeological (and palaeontological) establishments

with a certain 'do-it-yourself' attitude; indeed, the present thesis is, to a certain extent, an example of such an attitude. Given the material requirements and the wish or lack of any other option but to 'keep it in the family', much stress has been laid upon the simplicity of analyses including their practicability without the need to ask for outside help. This attitude is well illustrated in the work of Laville:

> Unspectacular but effective, these [analytical] methods offer the advantage of great simplicity combined with ease of interpretation. (1976a:154)

> L'étude des phases argileuses par diffractométrie des rayons X n'a été réalisée que sur un petit nombre d'échantillons, l'utilisation systématique de cette méthode d'analyse imposant de solliciter trop souvent l'obligeance de spécialistes extérieurs à notre laboratoire. (1975:39-40)

> [Trans.: The study of clay mineral species by X-ray diffraction has been carried out on a small number of samples only, since the systematic use of this analytical method would require that we overtax the generosity of specialists outside our laboratory.]

Care should be taken not to seek too much independence. At the moment, 'specialists' usually accept the involvement of archaeologists:

> The archaeologist who ventures into this field will soon find himself at the frontiers of knowledge, probably asking questions to which there is at present no answer, and feeling tempted to digress into the fundamental petrological research which could provide answers. Such digressions should be encouraged, as they can only benefit both archaeology and sedimentary petrography. (Catt & Weir 1976:85)

However, if 'specialists' should come to feel that they are being bypassed or used merely as technicians for the more difficult analyses, they will certainly become much less co-operative. We, as archaeologists, simply do not have the knowledge, resources or man-power to function independently. If it is felt that a 'specialist' lacks understanding of an archaeological problem, it is hardly his fault. It is the archaeologist's job to explain the problem, to <u>make</u> the 'specialist' understand even if he appears to show little interest.

Simplicity is a valid objective but it may also be something of a trap. In any analysis, simple or complex, it is of paramount importance that the exact parameter being measured should be identified. The preliminary research that usually goes into the establishment of a complex technique will almost certainly have involved the clear definition of the parameter to be measured. However, many simple techniques are based upon empirical results, often derived from rather restricted study material. Other simple techniques may be developed, perhaps rather uncritically, in order to avoid a complex technique which apparently measures the same thing. The simplicity of the technique may often belie the complexity of the system under study.

In the following section, a detailed example is given of how a simple technique, developed in another discipline, has been applied uncritically in cave sedimentology in order to circumvent the need to carry out extremely complex analyses. The error is not at the interpretative level; it is at the level of the actual mechanism of the analysis. Quite bluntly, this technique does not measure what it is claimed to measure.

7.2. A Case Study: ΔpH and Clay Mineral Alteration

The ΔpH method of assessment of the degree of alteration of clay minerals has been used by some French sedimentologists in the cave situation (cf. Thibault 1968; Laville 1975).

Thibault states:

> Le ΔpH est la différence entre deux valeurs de pH obtenues sur deux suspensions du même échantillon, l'une dans l'eau distillée, l'autre dans une solution de KCl normal. La lecture qui émane de la mesure du pH avec KCl étant toujours plus faible, par suite d'un échange de bases entre les ions K^+ de la solution et les ions H^+ du complexe absorbant. (1968:2215)

[Trans.: ΔpH is the difference between two pH values obtained from two suspensions of the same sample, one in distilled water, the other in a normal solution of KCl. The reading derived from the pH measurement with KCl is always the lower of the two, due to base exchange between the K^+ ions of the solution and the H^+ ions of the adsorbing complex.]

Monnier, who has used the technique in several open sites, gives the theory in a little more detail:

> Les possibilités d'échanges $H^+ \leftrightarrows K^+$ s'accroissent avec l'ouverture des espaces interfoliaires des argiles. Les illites ont des feuillets étroitement joints par les ions K^+; peu de charges négatives restent disponibles pour retenir les ions échangeables. Par contre certains minéraux comme les Montmorillonites présentent une forte ouverture des feuillets et ainsi une capacité d'échange élevée. Les ions K^+ du chlorure de potassium s'échangent et libèrent des ions H^+ ce qui entraîne un abaissement du pH et une augmentation du ΔpH. La valeur du ΔpH sera d'autant plus grande que les argiles illitiques verront augmenter dans leur empilement le nombre de feuillets de type vermiculite ou montmorillonite. (1973:15)
>
> [Trans.: The possibilities of exchanges of the type $H^+ \leftrightarrows K^+$ increase with [wider] inter-plate spacing in the clay [minerals]. The illites have plates closely connected by K^+ ions; few negative charges remain available for holding exchangeable ions. Conversely, certain minerals like the montmorillonites have wide spacing between the plates and thus a high exchange capacity. K^+ ions of the potassium chloride are exchanged and liberate H^+ ions, causing a drop in pH and a rise in ΔpH. The value of the ΔpH will be greater, in proportion to the number of plates of vermiculite or montmorillonite type contained within the structure of illitic clays.]

To carry out this analysis, all that is needed in the way of equipment is a pH-meter, distilled water, potassium chloride and containers. Compared with X-ray diffraction, this simple method claims to give sedimentologists a cheap and rapid way of measuring clay mineral alteration. Brief references to the method have now appeared in English (Laville 1976; Shackley 1975) and its use will no doubt become widespread.

However, the underlying assumptions involved in this technique require further examination. The processes, cited above, that are thought to be responsible for the ΔpH phenomenon are the absorption of K^+ ions and the complementary desorption of H^+ ions, respectively to and from sites between silica tetrahedral layers in (2:1) clay minerals ("argiles illitiques", supra).

These sorption processes are two of a larger group which, together, may be referred to as the ion exchange capacity of the mineral. The total system is extremely complex: "The exchange reactions always proceed in a stoichiometrical manner [i.e. they are dependent upon the various chemical elements involved], but even so, there is no simple way to predict the ion exchange capacity of a given clay mineral."(Degens 1965:20).

Before further discussion of the theoretical chemistry of these reactions, it will be interesting to see whether the published ΔpH results themselves contain information which might suggest correlation with factors other than clay mineral type and degree of alteration.

Let us, for the moment, accept the above description of the ΔpH phenomenon. It is clear that, given a series of sediments with identical types of clay mineral but with varying quantities of those minerals, ΔpH should be directly proportional to the quantity, or percentage, of clay minerals present, granted that high potassium content is maintained in the liquid phase. Figures for the abundance of clay minerals in cave sediments are not available. Indeed, extremely costly X-ray diffraction analyses would be needed even to provide semi-quantification. In order to obtain a very coarse measure of clay mineral content, let us assume, just for the sake of argument, that this content is proportional to the total colloid content of a sediment.

Laville and others use material of <2mm particle diameter as samples for ΔpH measurement. Colloid content can be approximated by figures for material <2 microns. Actual values are only rarely cited in tables, so that figures have been derived as carefully as possible from graphed results unless otherwise stated.

First, Coupe [section] 3 of the Pech de l'Azé II (Dordogne) was chosen in order to investigate the possible correlation between ΔpH and colloid content (data from Laville 1975). A standard linear regression model was used, with n = 49, Y = ΔpH and X = %<2 microns. The resulting equation is:

$Y = o.5375 + 0.0108X$

$r = +0.5055$ (>99% significant; t-test)

(cf. fig.17)

By examining the ratio between the deviations mean square and the original mean square,

$$\frac{[\Sigma(y - \bar{y})^2 - \Sigma(\hat{y} - \bar{\hat{y}})^2]/(n - 2)}{\Sigma(y - \bar{y})^2/(n - 1)}$$

where $\bar{y}$ = mean y

$\hat{y}$ = y value derived from regression equation

$\bar{\hat{y}}$ = mean derived y

- it can be seen that the <2 micron content accounts for c.29% of the variance in the original ΔpH values.

At Caminade-Est (Dordogne) (Thibault 1968; Laville 1975), the results are:

$Y = 0.6682 + 0.0179X$ (n = 20)

$r = +0.5532$ (>95% significant)

(cf. fig.18)

Approximately 27% of the variance in the original ΔpH values is accounted for by <2 micron content.

The site of Les Jambes (Dordogne) is again quoted by both Thibault, with figures in tabular form, and by Laville, although Laville's ΔpH graph appears to have the scale incorrectly displaced by +0.1 unit. Even here, with a sample of only 8, the results are:

$Y = 0.1280 + 0.0169X$

$r = +0.8284$ (>99% significant)

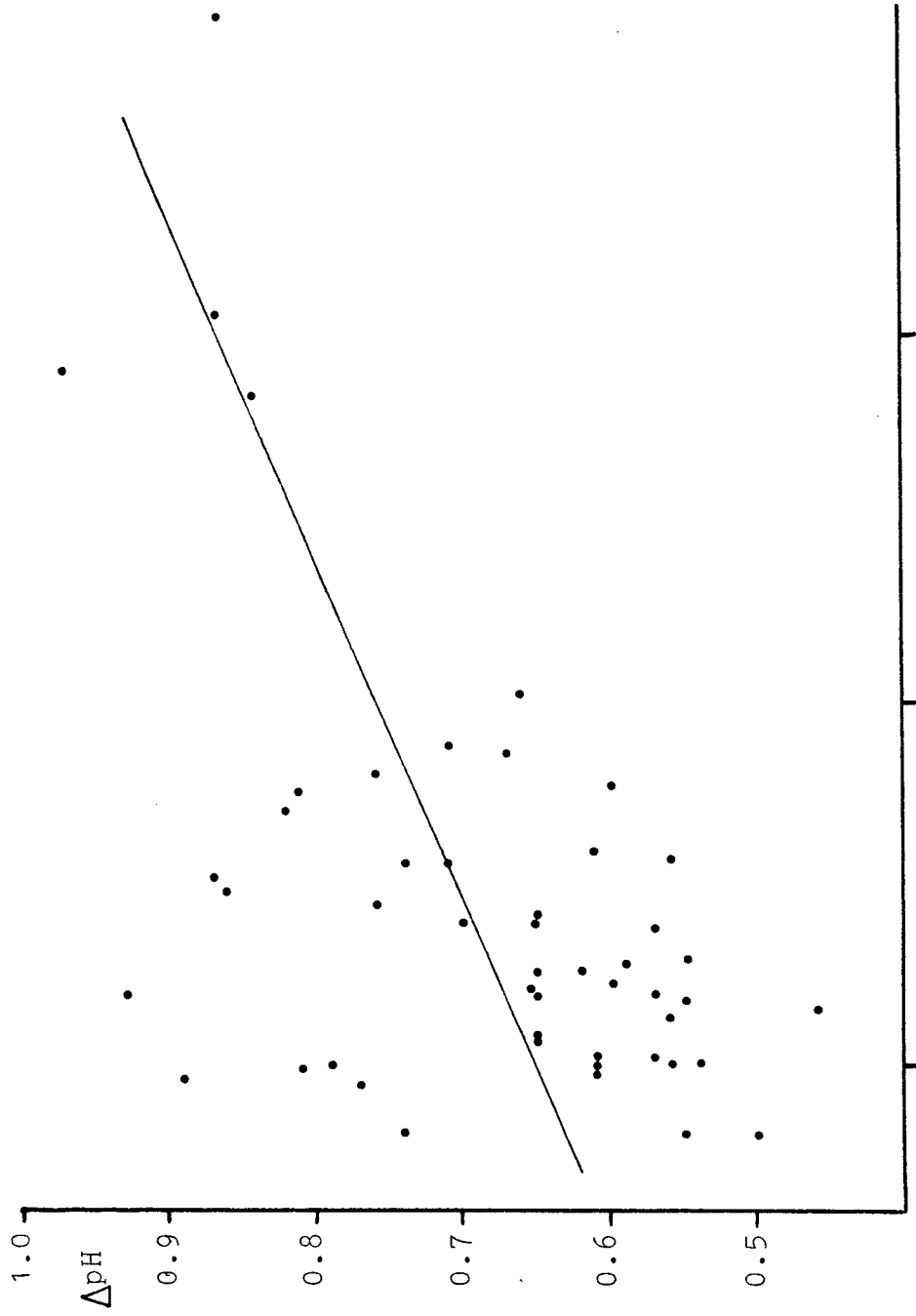


Fig.17 Delta-pH/colloid Plot - Pech de l'Azé II, Section 3

(data from Laville 1975)

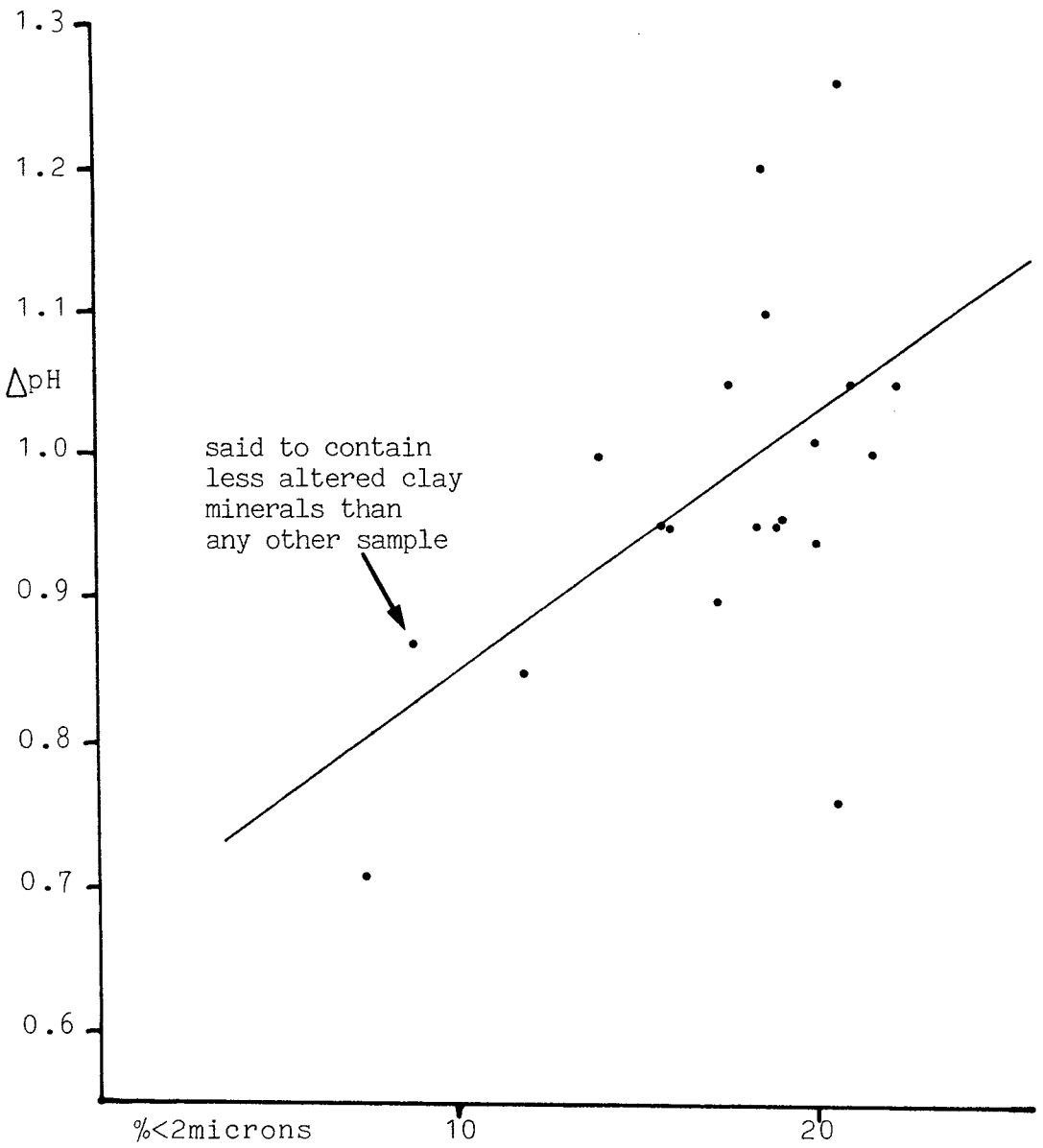


Fig.18 Delta-pH/colloid Plot - Caminade-Est

(data from Thibault 1968)

Figures (tabulated) from an open-air calcareous site, Port-Morvan (Brittany) (Monnier 1973), were also checked to see how general this effect might be. With n = 18, the correlation coefficient r was found to be +0.7693 (>99% significant) (cf. fig.19).

The proposition that ΔpH should be correlated with clay mineral content (quantitative) would seem to be supported by these results. Note, however, that the statistics were used only as an informal tool. Clay mineral content was not measured and the criteria of randomness and normality, required by the tests, were not necessarily met by the data. Indeed, theoretical considerations would suggest that a logarithmic rather than a linear model would be more appropriate (infra); such a model was not used because there were already too many uncertainties ever to hope that an exact result might be achieved. Such results as have been produced are suggestive, but they must not be used to 'recalibrate' the original ΔpH values.

Laville (1975) makes no comment whatsoever concerning the effect and uses the raw ΔpH values for direct interpretation. Thibault, when considering Caminade-Est, notes:

> La représentation graphique du ΔpH . . . offre une nouvelle fois de curieuses similitudes avec celles des colloïdes. (1968:2216)
>
> [Trans.: The graphic representation of ΔpH . . . offers once again curious similarities with that showing the [proportions of] colloids.]

In his general conclusion to this article, he states:

> Lorsque, dans les profils, les quantités de colloïdes, et par conséquent d'argiles, s'accroissent, il n'y a pas forcément modification du ΔpH dans le même sens. (ibid., p.2218)
>
> [Trans.: When, in given sequences, the quantities of colloids, and therefore of clay [minerals], increase, there is not necessarily a corresponding increase in ΔpH.]

Thibault is quite right in this last statement but if, as our results suggest, clay mineral content could account for

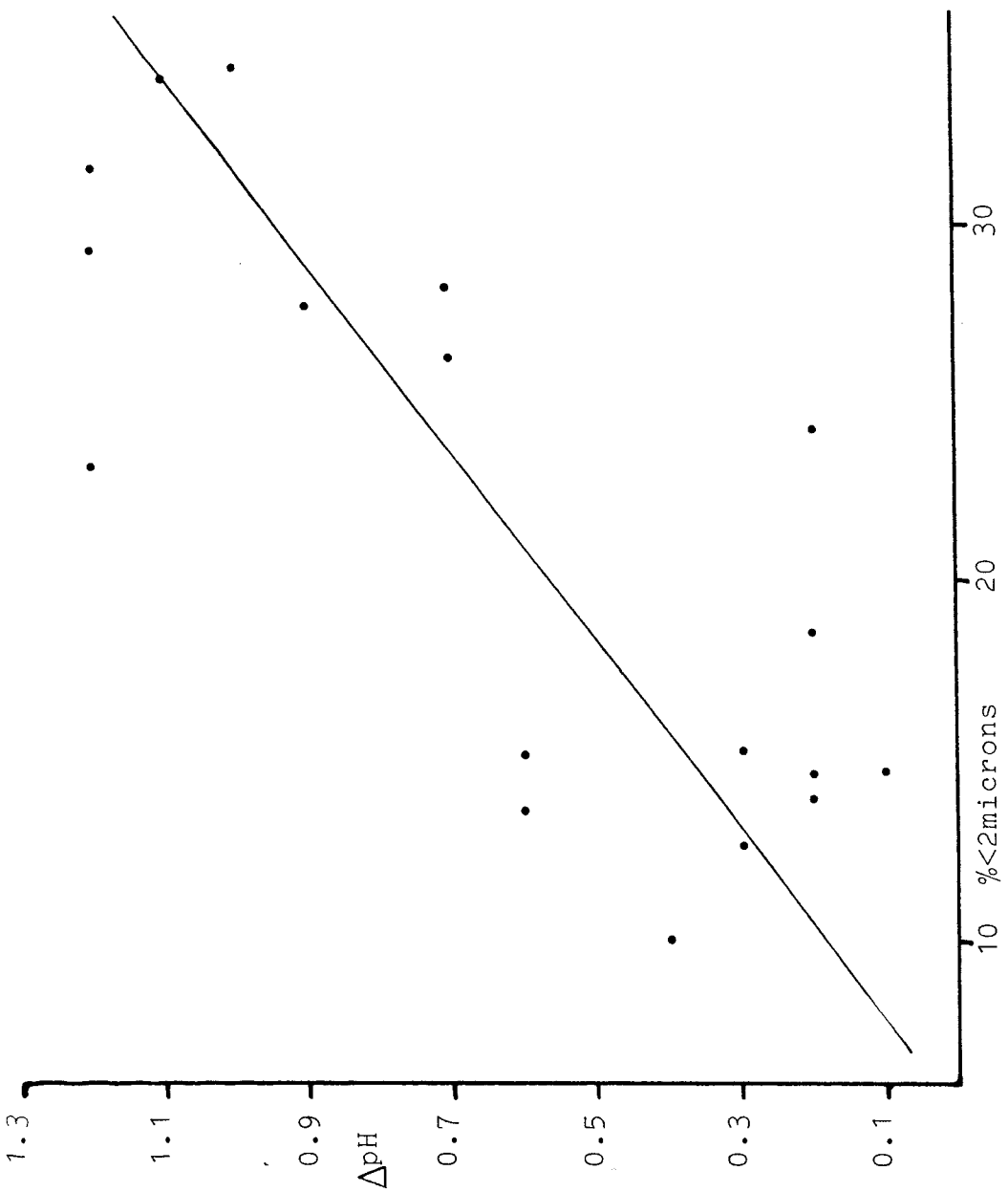


Fig.19 Delta-pH/colloid Plot - Port-Morvan

(data from Monnier 1973)

something in the region of 25% of the variance in ΔpH values, the effect can hardly be ignored.

So far, only archaeological sites have been discussed. The ΔpH technique was in fact developed by researchers interested primarily in agronomy (Franc de Ferrière et al. 1963). In this paper there is a diagram, reproduced here as fig.20. Clay mineral identification was by X-ray diffraction. These authors state:

> Dans les parties basique et faiblement acide du graphique, et pour les trois types de sols, le ΔpH croît assez rapidement avec la diminution du pH. Ceci est dû au fait qu'il y a échange des ions K^+ du KCl, d'abord avec les cations alcalino-terreux, puis, au fur et à mesure de la diminution du pH, avec un nombre de plus en plus grand d'ions H^+. Leur apparition entraîne un abaissement progressif du pH, déterminant une hausse correspondante du ΔpH.
> Dans la partie fortement acide du graphique, les ions H^+ sont naturellement présents dans les sols à une concentration qui s'élève avec une extrême rapidité en progression logarithmique, en fonction de la diminution du pH de 5 à 4. L'influence relative des ions H^+ libérés en nombre restrient par échange est faible. Ceci entraîne un abaissement très rapide du ΔpH. (1963:4941)

> [Trans.: In the basic [alkaline] and weakly acid areas of the graph, and for the three types of soils [not counting the 'interstratified' type; cf. fig.20], ΔpH rises quite rapidly with dropping pH. This is due to the fact that there is exchange between the K^+ ions of the KCl, first with alkaline earth cations [eg. Ca^{+2}, Mg^{+2}], then, as the pH drops, with a larger and larger number of H^+ ions. Their [H^+] appearance causes a progressive lowering of the pH, giving a corresponding rise in the ΔpH. In the strongly acid area of the graph, H^+ ions are naturally present in the soils in concentrations that rise with extreme rapidity, in a logarithmic progression, as a function of the drop in pH from 5 to 4. The relative influence of H^+ ions liberated in small numbers by exchange is weak. This causes a very rapid drop in ΔpH.]

The relationship between pH and ΔpH is obviously curvilinear but, at the pH values above 6 which are likely to be found in most cave deposits, an approximate linear expression will give an order of magnitude. It would seem that, for a given clay mineral type, a rise in pH of one unit corresponds to a drop in ΔpH of c.0.4 units. The effect appears far from negligible. Note that the relationship between pH and ΔpH is expressed by Franc de Ferrière et al. in an empirical manner which rather obscures the underlying principle. The logarithmic progression in the number of H^+ ions

1 - montmorillonite
2 - interstratified illite/montmorillonite
3 - vermiculite
4 - illite

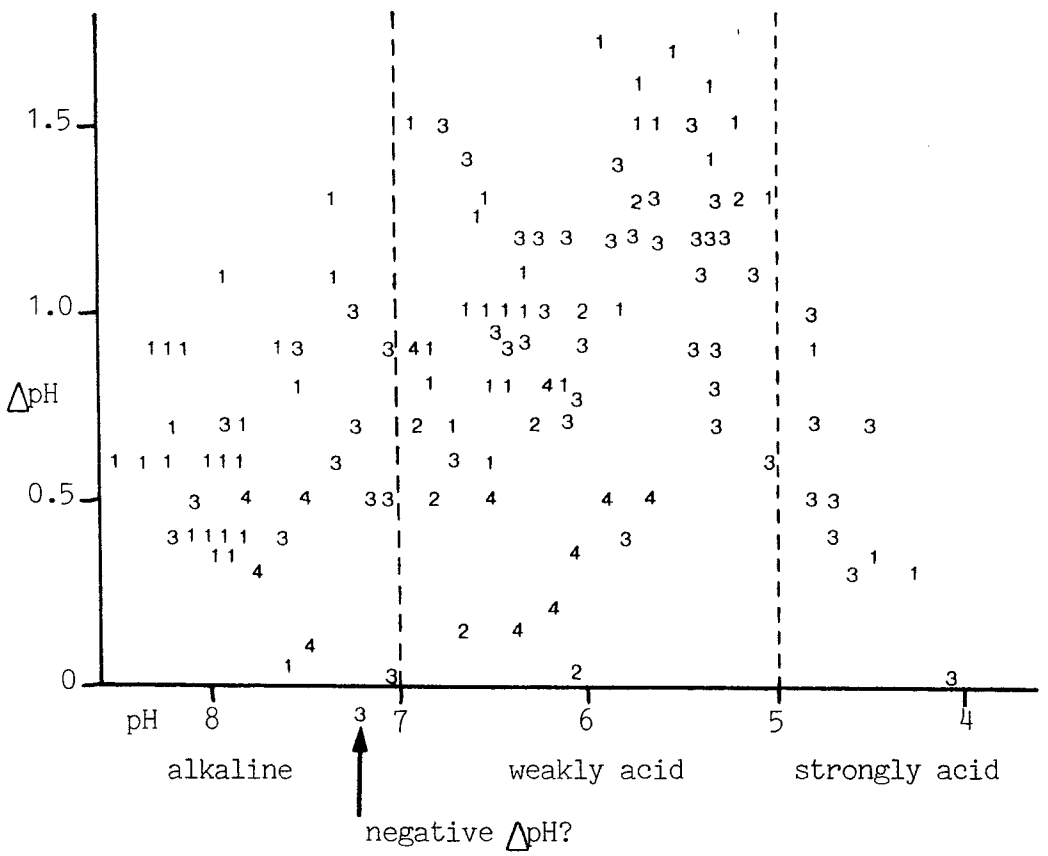


Fig.20 Delta-pH/pH Relationships

(after Franc de Ferrière et al. 1963)

is, of course, implicit in the actual definition of pH of any value: pH is the negative decimal logarithm of H^+ ion concentration in moles per litre. Thus, one litre of a theoretical solution of pH 0 contains 1g of hydrogen ions; of pH 1, 0.1g of hydrogen ions; of pH 2, 0.01g of hydrogen ions; etc.

Thibault (1968), in the article which introduced the technique to archaeological sedimentologists, supported the statement of Franc de Ferrière et al. concerning the role of pH and noted how his results showed a similar effect. The small number of results quoted in his paper are plotted here as fig.21. Laville (1975) makes no comment on this matter and does not quote his pH (H_2O) figures.

No further information can be extracted from the published accounts of ΔpH analyses. We may now turn to more general discussion of the geochemistry of clay minerals and other colloids in order to see whether our empirical criticisms are theoretically valid and also whether any additional factors may be expected to influence ΔpH.

Degens offers the following comments concerning ion exchange in clay minerals:

> (. . .) . . . we enumerate the following structural causes for the fixation of ions to clay minerals, and the simultaneous elimination from the liquid phase:
> (1) Unsatisfied valences produced by "broken bonds" at surfaces and edges of the clay particles
> (2) Unbalanced charges caused by isomorphous substitution
> (3) Dissociation of OH^- radicals, the H^+ of which may be exchanged
> (4) Accessibility of atoms in structural positions when brought to the exchange site as a result of changes in environment.
> Most significant among the environmental factors which influence ion exchange reactions are:
> (1) Availability of exchangeable constituents in the mobile phase
> (2) pH - Eh relationships
> (3) General chemistry of the environment
> (4) Pressure and temperature conditions.
> (1965:20-1)

First, let us note the second environmental factor quoted.

A - Les Jambes (illite + kaolinite slightly > interstratified montmorillonite)
B - Le Basté (illite >> clorite + chlorite/vermiculite)
C - Le Chemin Jean-Petit-de-Bas (no information)

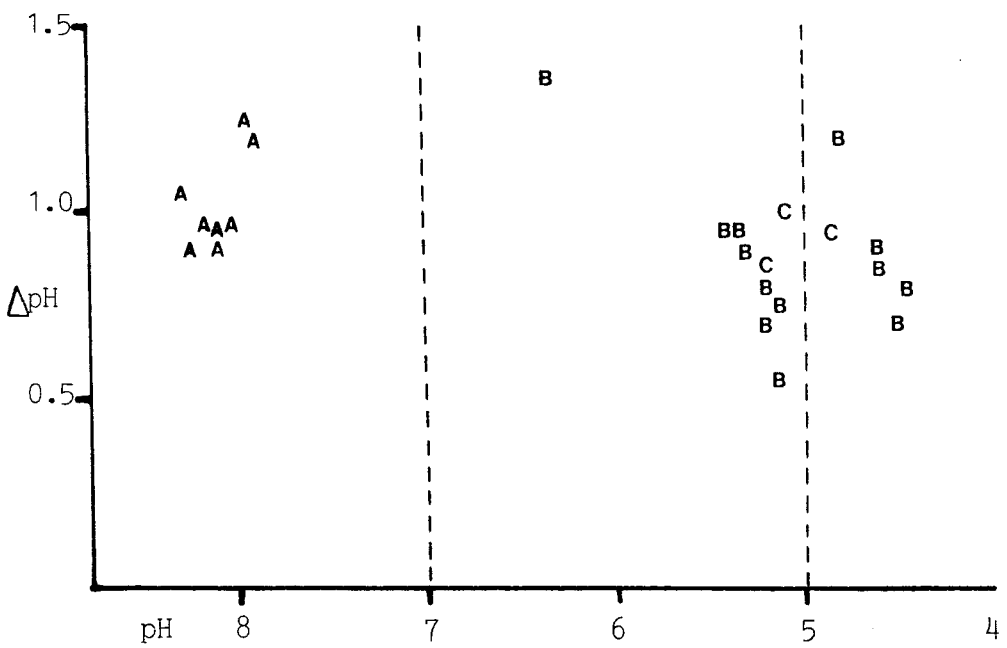


Fig.21 Delta-pH/pH Relationships

(data from Thibault 1968)

Eh is a measure of the oxidation-reduction potential of a substance, usually determined in a suspension of water. At an acidity level normally standardised at neutral (pH 7), this measure is used to express ion exchange capacity. We have noted that ΔpH is one aspect of the ion exchange capacity of clay minerals. Therefore Degens' proposition allows us to infer that ΔpH is pH-dependent. All theoretical texts consulted by the present author firmly support the empirical findings of Franc de Ferrière et al.

The published accounts of ΔpH mention only substitution at sites between the silica tetrahedral layers in (2:1) clay minerals. However, it is clear from Degens' list of structural causes for fixation that other sites will also be available. A theoretically perfect clay mineral would be an infinitely large, and endlessly repetitive, three-dimensional object. Of course, this is not so in practice; in two planes, the sheets are broken ("edges") and, in the third, the pattern of sheets fails to be repeated ("surfaces"), thus defining a finite object. Edges and surfaces are chemically reactive sites available for ion exchange; all the causes listed may operate at these sites. This suggests that different clay minerals, (1:1) as well as (2:1) types, will receive different proportions of their total ion exchange capacity at interior or exterior sites. Thus kaolinite receives its most important contribution to exchange capacity from broken bonds at exterior sites, whilst the smectite group (including montmorillonite) receives less than 20% from this source (Yariv & Cross 1979). Note, in passing, that charges developed through isomorphous substitution (substitution by an ion of similar atomic or molecular radius, but different valency), which is more common in the complex minerals such as smectites, are not affected by environmental pH. This variability in the relative importance of interior and

exterior sites is not particularly critical when it comes to differentiating between more or less pure samples of, say, kaolinite (a (1:1) type) and illite (a rigid (2:1) type), on the one hand, and smectite (an expanding (2:1) type), on the other. The following 'representative' figures for total ion exchange capacity (expressed in milliequivalents per 100g) of various clay minerals show how the complex types, with very active internal and external sites, are usually well differentiated from the simple types, with mostly external contributions:

kaolinite	1 - 10	simple
illite	10 - 40	
smectite	80 - 140	complex
vermiculite	100 - 180	

(cf. Brownlow 1979)

Consider, however, how particle size might have influenced the above figures. Clay minerals ". . . are of fine grain, generally less than 5 microns in size, commonly even smaller, and in some cases as small as a millimicron."(Pettijohn 1975:266). In fact, some clay minerals may be as large as 50 microns. Different clay minerals are usually found in different parts of this size range; for instance, illite particles (mean commonly 0.1 to 2.0 microns) are often larger than smectite particles (mean commonly 0.05 to 0.50 microns). Even so, a single type may be present at sizes that differ by factors of well over 1000. Smaller particles will have a higher 'external surface area/mass' ratio, that is, relatively more external sites will be available, not only through broken bonds but also through all the other listed causes. The commonly smaller size of smectites, compared with illites, is a major contributing factor to their generally higher ion exchange capacity.

Now consider a sequence of sediments with a uniformly high

clay mineral content in the colloids of all layers, thus ensuring relatively high absolute ΔpH values. The clay minerals of each layer are of identical type and degree of alteration. However, each layer has a different mean and spread of clay mineral particle size. The variation in ΔpH, or in total ion exchange capacity, would be largely a function of particle size. If the clay mineral involved was a kaolinite, with its most important contribution to ion exchange from broken bonds, the degree of variation would be high. On the other hand, if the mineral was a smectite, with a significant contribution from interior sites the availability of which is less affected by particle size, the variation might be slightly dampened. With respect to this question of particle size, if pure 'colloids' or 'clays' were used as ΔpH samples, the choice of the upper size limit of this class, which varies from 5 microns for the U.S. Bureau of Soils to 1 micron for many schools of geology and geochemistry, would obviously influence the results, depending upon which clay minerals were present and upon the particle size range of those minerals. Most importantly, publication of the limit used would not allow the reader to adjust the results for comparison with work undertaken using a different limit.

The first environmental factor listed by Degens (supra) concerns the availability of exchangeable ions in the liquid phase. In the case of ΔpH, a massive imbalance between the solid and liquid phases has been induced by the introduction of the potassium chloride. Cation exchange, between K^+ and H^+, is therefore guaranteed. But there is also a superabundance of Cl^- ions. Anion exchange, between Cl^- ions in the liquid phase and OH^- radicals in the solid phase, is also theoretically possible. Through proton dissociation (supra) of the OH^- radicals once

liberated, H^+ ions may become available in the solution, thus lowering pH. K^+ ions may exchange directly with H^+ ions derived from dissociation of OH^- radicals at the clay mineral boundaries, but the route via anion exchange will have a different reaction rate. Information concerning anion exchange in clay minerals is scarce and its effect on ΔpH values is at present unpredictable. However, it should be noted that such an effect is likely to be pH-dependent, since the contribution to the total negative charge on a clay mineral of proton dissociation at its surface is known to increase with rising pH (Yariv & Cross 1979). Adsorption of Cl^- ions ceases altogether above pH 8.

Clay minerals are usually discussed as crystalline substances, since their chemistry is easier to comprehend when they are considered as regular forms. There will almost certainly be irregularities or distortions in the lattice structure which govern the degree of crystallinity of a given mineral. Such distortions, which result in a higher ion exchange capacity, are not necessarily a function of any meaningful alteration but may be the result of the environment of formation of, say, a volcanic rock far removed in time from our cave sediment. Truly amorphous alumino-silicates, such as allophane, often have a very high cation exchange capacity and a not inconsiderable anion exchange capacity, both of which are pH-dependent. The behaviour of such non-crystalline substances is impossible to predict.

It has frequently been stated in the above arguments that not all colloids are clay minerals. Some substances, such as relatively pure silica colloids, will not enter into the reactions involved in the ΔpH phenomenon, although their presence will 'dilute' the exchange effect. Other substances cannot be dismissed so lightly.

Organic colloids may act as anions (micelles) to adsorb cations at an extremely high capacity, in some cases far in excess of even the capacity of smectites. The negative charge is mostly produced by carboxylic (-COOH) and phenolic (-OH) groups at the micelle surface, although chemisorption within the micelle structure itself may also occur. It is very difficult to measure the true quantity of organic material in a sediment, since such a wide range of substances are grouped under this heading. Only the humic and fulvic acid components are usually estimated, by means of solution in alkalis. Humin, polysaccharides and fossil hydrocarbons may all be present in sediments in large quantities and they will certainly take part in ion exchange. Note that organic colloids may themselves be adsorbed onto clay minerals, both at external and internal sites, and at various bond strengths. Even if extensive measures are taken to remove organics before ΔpH determination, and if this could be done without disturbing the clay mineral structure (infra), some adsorbed organics would almost certainly remain, adding their exchange capacity to the total ΔpH value.

If either organics or carbonates are destroyed in preparation for ΔpH analysis, carbon dioxide will become available for inclusion in sorption reactions. Such minerals as olivines will adsorb H^+ ions in the presence of CO_2, liberating metallic ions; this will change the hydrogen availability during the ΔpH measurement. The serpentine and zeolite groups have a high ion exchange capacity in alkaline conditions. Micas, especially if they are altered or mechanically damaged and comminuted, will adsorb K^+ ions, not necessarily with the release of H^+ ions. Hydrated iron and manganese oxides have low ion exchange capacities but a significant contribution might be made if the oxides were

present in sufficient quantities. Some of the above reactions are very slow compared with exchange involving clay minerals and organic micelles; the latter are effectively instantaneous once physical proximity of exchangeable ions has been achieved. To exhaust the potential exchange capacity in a closed system involving micas may take several weeks. Nevertheless, since 24 hours in suspension are usually allowed for ΔpH measurements (Thibault 1968; Monnier 1973), such slow reactions might affect the results. If particles larger than colloids are present in the sample, ion exchange involving silicates other than clay minerals will become increasingly significant. Note also that sorption reactions are steadily more pH-dependent as particle size increases.

Although a specialist geochemist would no doubt be able to give more precise details, the above discussion provides some idea of how individual minerals and other substances might react in the ion exchange situation. However, it would appear extremely difficult to foresee how a total real system might behave. Many substances, such as the complex phosphates, have not been considered. For instance, the large $(H_2PO_4)^-$ radical should be exchangeable for OH^- radicals. How might the many reactions noted above affect each other? Are there any available catalysts or inhibitors that might change the reaction rates? The overall impression is one of extreme complexity, far beyond the possibility of evaluation in any practical situation.

Before leaving the subject, we should consider whether the physical techniques of ΔpH measurement might cause any variation in the results. Could flocculation of a sediment with a high colloid content physically inhibit ion exchange? Note that flocculation is pH-dependent. If we chose to use a mechanical

agitation technique to forestall flocculation, might not an abrasion-pH effect (Stevens & Carron 1948) influence the results, especially if particles larger than colloids were included in the sample? Removal of sediment from the cave environment will certainly induce chemical changes; pH values from measurements on site and in the laboratory are rarely identical. How might drying (necessary to provide two subsamples of equal weight) affect the results? Most colloids have weakly adsorbed water molecules at their surfaces but clay minerals, especially the expanding (2:1) types, may also have more strongly adsorbed water at interior sites. Organics, themselves with varying amounts of water, may also be present at these interior sites. Drying at different rates to different temperatures, or to different relative humidities (eg. in a desiccator), will change the chemistry and structure of clay minerals and organics to varying degrees. Even at 60°C, or during long periods of air-drying at low relative humidity (eg. in a normal laboratory), some irreversible changes are known to occur (cf. Mackenzie 1957). Given that we know so little about the reactions and reaction rates involved in ΔpH analysis, how can we assess the importance of fluctuation, during the measurement, in the ambient temperature, a factor which must have some effect on all chemical reactions by very definition?

Finally, we might point out that understanding of a technique would be best accomplished if all results, no matter how aberrant they may seem, were published. Franc de Ferrière et al. state: "Le pH mesuré avec KCl est toujours plus faible qu'avec l'eau distillée. [Trans.: pH measured in KCl is always lower than in distilled water.]"(1963:4929). Sedimentologists working in caves have repeated this statement (cf. Thibault 1968:2215, quoted above). Some of the reactions noted in the

present discussion could result in the pH in KCl being higher than that in water. This author has been present when such results did in fact occur. Some error in manipulation was usually thought to account for the departure from received theory. Sadly, no negative ΔpH values have ever been reported, save perhaps for a single uncommented point on the graph of Franc de Ferrière et al. (cf. fig.20 in the present work).

The ΔpH method has been examined in detail since it affords a thorough example of how sedimentologists may be tempted to use a simple technique, without proper consideration of the complex and unpredictable system which the technique claims to 'measure'. The end result of much that has been mentioned above cannot as yet be evaluated, but the mere fact that clay mineral content itself cannot be accurately quantified is sufficient to cast doubt upon the validity of the technique. For his part, there is no conceivable situation in which the present author would feel justified in measuring ΔpH (cf. Collcutt 1979a).

If a measure of ion exchange capacity is required as a very coarse approximation of the sort of clay minerals present, the author would be more inclined to recommend direct Eh determination (also available on most portable pH-meters). A whole sample can be used or, better still, the <45 micron fraction (the finest split that can be achieved using medium-priced sieves). Alkali-soluble organic content should also be measured and some idea of overall mineralogy would be helpful. This method too suffers from most of the drawbacks noted for the ΔpH technique. However, only one subsample is required and therefore it need not be dried before the Eh measurement, only afterwards. Furthermore, measurement is much more rapid, which, apart from being efficient, reduces the probability of interference from slow ion exchange

reactions. Measurements can easily be carried out on site with very little disruption of the sediment's natural environment. An added bonus is that Eh data may be available from previous soil and geological surveys in the area of the cave, whilst ΔpH data almost certainly will not.

The ΔpH method has been criticised at the level of the mechanisms of the analysis. However, even allowing that an estimate of ion exchange capacity if provided by this method, the normal interpretation of the results is wholly unjustified. The reasoning is that high ΔpH values equal high exchange capacity, equals certain types of clay minerals, equal alteration of the whole sediment, equals 'weathering/pedogenesis', equals climatic amelioration. This reasoning is unsound at every single stage. The argument that relative differences in the value of ΔpH in different sediments indicate relative differences in climate is even more absurd. For further discussion of clay minerals, see chapter 12.

8. THE RECOGNITION OF THE EFFECTS OF ICE

8.1. Ice in the European Quaternary

The greater part of the Quaternary Period consists of the Pleistocene Epoch, a division of geological time first defined by Lyell on palaeontological grounds. However, due to a growing recognition of the impressive effects of glaciers, this epoch soon became known as the Ice Age. Glacial and related (periglacial) deposits do indeed constitute the bulk of pleistocene sediments in many parts of Europe, even though 'cold' climates as such operated for considerably less than half of the complete time span. Indeed, the importance of ice is no longer thought of as being restricted to the Pleistocene and geologists now speak of the Late Cainozoic Ice Ages, running back as far as the Late Miocene.

> Although the Quaternary covers a much shorter period of time (about 2m.y.) than any other system, its deposits are preserved in a variety and wealth of detail unapproached by older periods. They embody a record of great climatic fluctuations which are the outstanding feature of the era [period] and it has long been accepted in most parts of the world that these climatic cycles are the natural basis for time sub-divisions. [...] The Quaternary Era [Period] Sub-Committee has therefore accepted climatic fluctuation as the guiding principle for defining the stages of the British Quaternary. (Mitchell et al. 1973)

The Quaternary, then, is a period comprising many climatic fluctuations between approximately present-day conditions and colder conditions. At least in Europe, the larger part of the lithostratigraphic record will comprise sediments developed under colder conditions, simply because terrestrial sedimentation

rates are often orders of magnitude greater under such conditions. As a general proposition, it seems likely that this may be true for cave as well as open-air deposits. Chronostratigraphic units of the rank of stages (with their geochronologic equivalents, ages) are defined in terms of climatic units (glacials and interglacials for the later Pleistocene). Usage for finer subdivisions has not yet been standardised so that substages (and subages) and chronozones (and chrons) may apply to a variety of minor time units. There is a tendency for even these minor units to be defined in terms of climate.

The present author has no great objection to any of the above points. Despite the fact that a climatic basis for time units often leads to major problems in practice, it is difficult to see what other currently available criterion could be used that would result in fewer problems. However, this reasoning concerning time units has had a curious effect upon quaternary geology in general. Because climatic fluctuations are known to have occurred, observed sedimentary phenomena must therefore be due to such fluctuations and must be interpreted in terms of such fluctuations. This is climatic determinism. Moreover, because the major part of the terrestrial lithostratigraphic record comprises sediments that were probably laid down under 'cold' conditions, European quaternary geology at its worst may reduce to the simple question: 'How much ice was there around at this time?'. We may call this 'cryodeterminism'. Seen from the present, the European Quaternary was (is?) a very chilly time.

The author in no way disputes the importance of cold climates. Many sedimentary phenomena may indeed be due, more or less directly, to cold climates. However, there is a very simple

way to test the validity of 'cryodeterminism' as a practical procedure and that is to suggest that its proponents prove it in specific instances. Given this requirement, the European Quaternary, in caves at least, warms up considerably and a great number of the climatic 'wiggles' previously defined just thaw away. Of course, this 'temperate' Quaternary is no nearer the truth of the matter than is the 'arctic' Quaternary usually recognised in caves. When we observe cave sediments, we know that there is a good chance that they were developed under cold conditions, but this is no reason to use facile arguments to link the observations with the effects of ice. In this chapter, it is argued that very few unequivocal indicators of the activity of ice are known to exist in European cave and shelter deposits. This may be because ice was not very common in caves - a proposition which the author does not in fact think is very likely - or because we simply have not found adequate methods of recognising phenomena certainly produced by ice. The way ahead is therefore clear.

The first step is to stop thinking of ice as a climatic phenomenon. It is a meteorological phenomenon, active in just about every part of Europe now. It is a medium, just like water or air, which is involved in a whole range of sedimentary processes. Some of these processes, such as glaciation or permafrost formation, are much more restricted areally now than during some parts of the Quaternary. Others, such as cryoclastis, are still operative over wide areas and it is the intensity of the process which is important.

Ice acts in five main ways: as a mass erosive agent; as a mass transporting agent; as an energy converter; as a physical restriction; and as a water source.

The first two roles are restricted to glaciers (sensu stricto) and shore ice. Primary glacial deposits may occur in shelters and wider cave entrances (which will therefore tend to be hidden under till), but the physics of ice flow inhibits the penetration of glacier ice very far into a normal cave system (cf. Ford 1976). Glacial deposits will most commonly be found in caves in a derived state. Deposits and erosion forms due to lake- or seashore ice have not been described in the cave context, to the author's knowledge.

When water freezes there is an energy flux. Some of the energy is lost but some is absorbed in the crystallisation process, in terms of both an overall 9.08% expansion and an oriented force associated with differential growth along specific crystal axes. Freezing therefore converts heat into potential pressure (stress). Bridgeman (1912) calculated that, as pure water is cooled to -22°C (the point at which ice changes to another form) in a completely closed space, a theoretical maximum pressure of 2100kg/cm² (i.e. over 2000 atmospheres) is produced. In practice, such a pressure is never reached because the water is never completely contained and it always has some impurities (dissolved gases and salts). Nevertheless, pressures capable of overcoming the tensile strength of many rock types are certainly produced (cf. Everett 1961; Mellor 1971). It is this role of ice that will be the main focus of the present chapter, first with respect to relatively uncontained effects and then to relatively contained effects.

Ice or frozen sediment may act as a physical restriction. This is most important with regard to water circulation. Depending upon the perspective, water can either be kept out (eg. sealing of a cave system by a glacier or by permafrost)

or it can be kept in (eg. water build-up in sediments with underlying and/or overlying ice). When ice (either ground ice or snow) melts, large amounts of water often become available rather quickly, an effect that more than compensates, in geomorphological and sedimentological terms, for the fact that most cold regions, especially at low altitude, have comparatively low precipitation. These two roles of ice have important dynamic effects in as much as they control water availability, a subject which will recur throughout this text.

5.2. Ice in Sediments

As was stated in section 2.2., practically nothing is known about contemporary ice in cave, or even shelter, sediments. The author has noted only weak, superficial (<5cm) freezing in sediments near limestone rockfaces or barely within cave entrances at various sites in Britain, when more exposed (and finer grained) sediments in the vicinity were experiencing at least twice this penetration. Also, there is a tendency for snow banks to melt very quickly near rockfaces, so that there will often be a patch of bare ground between the remaining snow and the rock of one or two metres. However, useful generalisations can hardly be extracted from such casual observations. The author has never had the opportunity to examine cave sediments under really cold conditions for any significant length of time and he has never come across a published report of such an examination. Ice must form to some extent, at least in some shelter deposits, but our understanding will obviously remain severely restricted until an actualistic data base can be established. For the time being, a theoretical approach will be

necessary, drawing upon the considerable body of information available from open sites, in the hope of identifying unequivocal structural features which we might expect to find in ancient cave deposits. At the same time, we must bear in mind the peculiar properties of cave sites which may affect the formation of such features.

In this section, we are concerned with the effects of freezing in uncontained situations, that is, within the bulk of a loose sediment body. It is clear that, at most natural rates of temperature decline, the mere fact of a single freezing episode may have practically no effect on a sediment whatsoever. It is the processes of repeated freeze-thaw and ice segregation which produce significant changes within a sediment.

For freezing conditions to have any great effect, a sediment must contain water. Although general availability of water in the environment is obviously involved, a more important factor is sediment permeability. In a coarse, permeable sediment, water will drain away quickly and the 'average' water content (the water which is retained by surface tension) available for ice formation will be low. In a fine, less permeable sediment, the 'average' water content will be higher. This model is made much more complex by two additional factors: (1) water migrates through a sediment during freezing episodes (ice segregation), and (2) water will not necessarily freeze at 0°C.

As air temperature drops, sub-zero conditions develop from the ground surface downwards into the sediments as a freezing front, which will be roughly horizontal (parallel to the ground surface). Ice will begin to form from the water contained within the pores of the sediment. Once freezing has

started, water in the underlying unfrozen sediment will migrate to the freezing front. There are two theories that have been proposed to account for this phenomenon, both of them depending upon the fact that water is a dipole that forms a charged double layer of molecules on the surface of a mineral (in this case ice). Chalmers and Jackson (1970) argue that there is an energy gradient, decreasing from free water, to the double layer, to the ice. This gradient causes water to migrate to the freezing front. Cass and Miller (1959) argue that there will be a build-up of cations in the double layer, causing purer water to be drawn to the freezing front by osmosis. Whatever the exact process, water certainly does move and ice crystals build up at right angles to the freezing front. As long as pore water is constantly available at high pressure, the front will not advance appreciably deeper into the sediment. Most impurities (either particulate or in solution) will also be segregated, so that the ice will be relatively pure. Segregation ice obviously takes up more volume than the original pore space, so that overall expansion of the deposit will result.

Water is present in a sediment at various bonding energies (cf. p.366). Some water adsorbed onto clay minerals will not become available until very low temperatures (in excess of -20°C) have prevailed over long periods, and only then if the sediment is otherwise very low in water vapour. More water is less strongly attached to fine particles. Water in general will be subject to surface (capilliary) tension, which may cause it to migrate more slowly or more rapidly, depending upon the size and configuration of pore spaces and on pore water pressure. Solutes in water will usually depress the freezing point below zero. All these factors result in the fact that, at a given

temperature and after a sufficient time, nearly all the water in a coarse, permeable sediment will be frozen, whilst a fine, less permeable sediment may still have considerable amounts of liquid water (20% and more of dry sediment weight at -20°C in some cases). Stated another way, the water in a coarse, permeable sediment will freeze in situ rather quickly, without much ice segregation, whilst water will remain available for ice segregation in fine, less permeable sediments for long periods after the onset of freezing.

An additional factor is whether or not a sediment is a closed system. A limit to both pore and segregation ice is set by the water content of the sediment. In a closed system, the sediment is quickly dewatered (or even truly desiccated) and no furtner ice can form. However, in an open system with a significant water reservoir beyond the reach of freezing conditions, huge quantities of water may be drawn to the freezing front to produce ice bodies many metres thick. Freezing of part of the pore water in a coarse sediment produces cryostatic pressure that may inject some of the remaining water into a neighbouring fine sediment, there to be utilised in additional ice segregation.

The subject of ice formation in sediments is extremely complex; more detailed discussion and a comprehensive bibliography may be found in Washburn (1979). Various 'rules of thumb' exist which may be some help in the decision as to whether or not a sediment is likely to be 'frost-susceptible'. Note that, in the context of loose sediments, frost-susceptibility is the propensity for hosting segregation ice and thus for sediment displacement (heaving, thrusting, sorting, etc.). In a hard rock, frost-susceptibility is the same basic property but

it is expressed in terms of the likelihood that the rock will fail during freezing episodes (cf. section 8.3.). In general, a sediment will not be frost-susceptible at all if it contains less than 3% of material finer than 0.02mm, although otherwise well sorted sands and gravels may have up to 10% before starting to become susceptible (Casagrande 1931). The mineralogy of the fine material is important with respect to its electrochemical properties. Taber (1930) found that quartz dust, with a maximum diameter of 0.07mm, was not susceptible, and Williams (1957) reports a similar result for morainic sediments with c.6% of material less than 0.02mm and c.15% less than 0.07mm. Linell and Kaplar (1959) state that kaolinite induces susceptibility at lower proportions than do illites and montmorillonite. Fabric may have some effect and Taber (1930) reports greater susceptibility in sediments with horizontal, tabular particles. Kaplar (1974) reports rate of heave (roughly equal to the rate of ice segregation) for a whole series of sediments, of different textures, tested in the laboratory, but the results are only comparative and the actual vertical displacement cannot be used for predictive purposes in real situations. Linell and Tedrow (1981) give further practical information on this subject, including the observation that, with coarser sediments, susceptibility increases with increasing compaction (closer packing leading to a finer pore network). All authorities are agreed that the most frost-susceptible sediments are silts (especially fine to medium silts) with a small proportion of colloidal material (inorganic or organic). As long as there is a more or less continuous matrix of such sediment, quite large proportions of coarse sand or gravel can be included as floating elements without drastically reducing the ice segregation

potential.

From the above generalities, it would appear that many cave sediments should be at least moderately frost-susceptible. Badly sorted deposits, with appreciable silt and colloid contents, are very common in British caves. However, at least in sediments near an entrance, compaction is often extremely low because the deposit is usually supported by coarse limestone clasts, which themselves are seldom arranged to give anything like maximum packing. The pore network is therefore coarse, often including major airspaces, and bulk particle size distribution may be a far from adequate indicator of frost-susceptibility. More compacted sediments are usually associated with processes originally involving more than average quantities of water (eg. fast mass movement, stream flow, etc.); such sediments may often be very dense and contain large proportions of colloids, thus reducing susceptibility because of very low permeability.

The availability of water is also a problem. Near a cave entrance, sediments are often comparatively dry, even ignoring the effect of the coarse pore network. Water tends to be derived overland from bare rock catchments near and above the cave, with a very fast response rate. At the dripline, or below larger fissures, water becomes available in quantities sometimes many times greater than the local, open surface precipitation, but this input is usually rather sporadic. When water penetrates the sediments, the rapid through-flow may cause piping (removal of fines from relatively wide sub-surface drainage channels) as well as the sealing of peripheral pores by carbonates. Thus, the pattern of drainage within the sediment becomes increasingly structured, which should reduce frost-susceptibility even further. The author has observed

occasions when the surface of deposits was sodden with water from drips and wash but, 5cm below the surface, the sediments were quite dry, except for specific points in cut sections where water was trickling out of preferential drainage pipes. Deeper into caves, the sediments usually become increasingly damp, and remain so even on cut sections. Water availability is much more homogeneous, a fact which encourages the gradual filling of voids between coarse particles and an overall rise in bulk density. These sediments are usually more compacted in the first place, because they are due to mass transport, various subaquaeous processes or other processes giving very slow sedimentation rates. Under the same temperature regime and assuming that sediments are not too impermeable, it might be expected that segregation ice would be more likely deeper into caves than at the entrance. On the other hand, sufficiently low temperatures will only occur deeper within a cave if there is significant air movement (cf. section 2.2.), and the temperature oscillations which are necessary for the accumulation of the sedimentary effects of segregation will be quickly dampened.

It may well be, therefore, that pore ice will be reasonably common in cave sediments but that significant segregation ice will be much less likely. Certainly, it has often been noted that, in better drained, stony , upland arctic soils, the ice content is much lower than normal, with little segregation ice (cf. Linell & Tedrow 1981:57). This suggestion is in urgent need of testing through a programme of observation in present-day cold climate caves.

The depth to which sediment is likely to freeze is modelled by various equations, usually involving estimates of

thermal conductivity of the frozen sediment, the ground freezing index (measured in temperature-time units, such as degree-days, for the period when ground temperature is below zero), latent heat effects, amount of water in the sediment, and the unit dry weight of the sediment (cf. Andersland & Anderson 1978). The difficulty with such formulae for cave sediments is the inhomogeneity of the deposits. The position of the freezing front is also influenced by the ability of water to migrate, as was noted above. The depth to which sediment is likely to thaw is given by similar formulae, using terms for the thermal conductivity of the thawed sediment and the thawing index (for the period when ground temperature is above zero). With very coarse approximations of the relevant parameters (derived from values in American Standard Units quoted in Washburn 1979, and Linell and Tedrow 1981), and using the formula of Yong and Warkentin (1975), the depth of freezing to be expected in a generalised (uncompacted) cave sediment, exposed to exterior meteorological conditions, might be:

$$x = \sqrt{\frac{48 k_f F}{L_w W y_d}}$$

Where x = depth of freezing (ft)

k_f = thermal conductivity of frozen sediment
= 0.8 BTU/ft/h/°F

F = freezing index (°F-days)

L_w = latent heat of freezing = 143.4 BTU/lb

W = water content = 0.1 (fraction of dry wt.)
= c.25% of saturation value

y_d = unit dry wt. of sediment = 100 lb/ft^3

With F at 3000 (cf. present-day Arkhangel'sk, very patchy

permafrost only, mean January temperature 10°F (-12.2°C)), depth of freezing would be c.9ft (c.2.7m); at 1500 (cf. present-day Norwegian coast at the Arctic Circle, no permafrost, m.J.t. 20°F (-6.7°C)), it would be c.6.3ft (c.1.9m); and at 500 (cf. present-day western Poland, no permafrost, m.J.t. 30°F (-1.1°C)), it would be c.3.7ft (1.1m). On the limestone of S.W. England, with F at c.20 today (m.J.t. 37°F (2.8°C)), freezing should penetrate to c.0.7ft (c.20cm), a figure in general agreement with the author's own casual observations. These figures are probably a slight over-estimation, since the effect of the favourable microclimate in some parts of caves (especially near or just outside the entrance) is not included (cf. section 2.2.). Nevertheless, it would appear that a large proportion of cave sediments in northern Europe would have been subject to freezing conditions at some time during their existence. Note that the same form of equation can be used to calculate the probable depth of thaw; thermal conductivity changes to c.1.2 and the thawing index replaces F in the formula. Given the suggested values of the other parameters, it can be seen that the freezing index needs to be c.50% greater than the thawing index before the first appearance of perennially frozen ground.

This brings us to a consideration of permafrost.

> Permanently frozen ground or permafrost is defined as a thickness of soil or other superficial deposit, or even of bedrock, at a variable depth below the surface of the earth in which a temperature below freezing has existed continually for a long time (from two to tens of thousands of years). Permanently frozen ground is defined exclusively on the basis of temperature, irrespective of texture, degree of induration, water content, or lithologic character. (Muller 1947:3)

This original definition clearly does not require that ice be present in the sediment.

> The term 'permafrost' is defined as a condition existing below ground surface, irrespective of texture, water content, or geological character, in which:

> a. The temperature in the material has remained below 0°C continuously for more than 2 years, and
> b. If pore water is present in the material a sufficiently high percentage is frozen to cement the mineral and organic particles. (Stearns 1966:1)

This definition is nearer the generally applied principle. However, it is not really explicit. Does the definition require that, if a small amount of water is present, it must be totally frozen, even if this does not suffice to solidify the sediment? Strictly speaking, 'permafrost' is therefore a sub-surface 'meteorological' concept and does not equate exactly with the presence of ground ice. Permafrost is often assumed to have been present during the Pleistocene if it can be argued that annual mean temperature was below zero.

It is ground ice that affects sediments, so that a vocabulary reflecting its occurrence is required. Since the term 'permafrost' is so commonly used to mean 'perennially ice-locked sediment', it will be used in this sense below. There would appear to be no advantage in an arbitrary distinction between 'permafrost' and 'pergelisol', and the two terms are considered to be synonymous. The upper surface at which ice ceases to be perennial is called the 'permafrost table'. Above this lies a zone with seasonal ice called the 'active layer' (or 'mollisol'). There are some situations in which a more or less permanent zone without ice (a 'talik') intervenes between the permafrost table and the active layer, but this would seem rather unlikely in the cave situation. Thus, in winter, the ground freezes more or less from the surface, with a 'freezing front' (supra) penetrating down through the active layer until it coallesces with the permafrost table. In summer, the ground thaws from the surface, the retreating level of ice being called the 'thawing front' (or 'frost table').

The permafrost table represents the lower limit to which the processes of the active layer may penetrate. The depth of an ancient active layer may therefore be estimated, in many situations, from the depth of certain fossil structures (infra). This is a useful concept in many open sites but, in cave sediments, it may be misleading. Very few structures of the requisite type have been identified in cave sediments, and, even if there were a good argument for the former presence of true permafrost, there is rarely any proof that those few structures could not have developed during the final degradation of the permafrost. Furthermore, allowance must be made for the shape of the freezing and thawing fronts. From the above discussion of ground ice, it is clear that these 'surfaces', were they to occur in caves, would be highly irregular, because textural parameters vary rapidly in both the vertical and horizontal planes. Indeed, columns, veins and discrete pockets and lenses of frozen material may be present at any stage of development (cf. p.271).

A number of impressive structures are associated with the activity of ground ice. These structures form a scalar hierarchy, with smaller scale structures combining in various ways to make up higher level structures. Unfortunately, many of the structural patterns that are really diagnostic of cold climates are either intrinsically on a large scale (eg. pingos) or are smaller structures repeated over wide areas (eg. various forms of patterned ground). Such clear indicators of ground ice will have little chance to develop in the areally restricted cave situation, even if all other conditions were favourable. The difficulty with the smaller scale structures is that virtually none of them, observed as isolated occurrences, are

restricted to permafrost zones. Indeed, many structural patterns, often found in association with permafrost, are not even restricted to seasonally frozen ground and may arise without the intervention of ice. Since the possibility of observing a range of structural phenomena, of the same age and over a significant area, will very rarely be available to the cave sedimentologist, the author seriously doubts that it will ever be possible to prove the former presence of permafrost in cave sediments, although a degree of probability might be achieved by proving stratigraphic links with extensive open-air deposits with unequivocal signs of permafrost. Our immediate task, itself far from easy, is to prove the former presence of ice, of whatever seasonality.

The structures of interest here may be divided into several groups: (1) barely macroscopic and microscopic structures; (2) various forms of convoluted bedding; (3) repositioning of coarse elements (sorting and orientation); (4) crack-and-fill structures; (5) patterned ground. These groups are not truly discrete and recognition of a structure depends largely upon perspective. Structures caused by mass lateral movement ('gelifluction') are discussed in the next chapter.

The very small scale structures associated with ice may prove to be very useful with respect to cave sediments, although very little work has yet been carried out in this field. Because geomorphologists and sedimentologists in general do not usually need to record such small structures, there being many larger structures available, only pedologists have taken any great interest (cf. Tedrow 1977). Fitzpatrick (1956) noted horizons in soils, which he called 'fragipans' because of their fissile properties. Tiny lenses of segregation ice,

forming at right angles to the direction of heat loss, tend to force particles to lie parallel to the lenses. In general, this will be a roughly horizontal orientation but other orientations may develop near irregular ice bodies. It was noted above that segregation ice forms more readily in parallel fabrics, so that ice and sediment are obviously part of a coupled system.

Goldberg (1979a), citing resemblances with fabrics described by Dumanski and St. Arnaud (1966), Romans and Robertson (1973) and van Vliet-Lanoë (1976), has suggested freezing as the cause of 'banded fabric' at the Pech-de-l'Azé II (Dordogne). This appears to be a sorting effect, whereby the plasma (silts and colloids) is concentrated into bands with little disruption of the supporting coarser grains. Goldberg notes that the fabric was visible on site as fine laminations, but neither he nor Laville (1973) gives a description of the geometry or persistence of this feature in any of its stratigraphic occurrences. Mücher and de Ploey (1977) figure rather similar fabric produced experimentally by wash on loessic sediments (cf. section 10.1.). The present author is not wholly convinced that Goldberg has succeeded in ruling out deposition by water. This underlines the importance of a range of perspectives; if microscopic observations do not form a continuum with clear and inclusive macroscopic description, there will always be an 'inference gap'. Note that, in fine sediments, movement (either on a small scale, as in injections, or mass movement), as well as wetting and drying, may induce similar microscopic structures, although persistent macroscopic laminations are unlikely.

As water freezes, impurities are segregated out, solutes being released either as gases or as solid precipitates. Gas bubbles may form a vesicular structure in fine sediments

(cf. Parea 1970; Bunting 1977), a proposition demonstrated experimentally by Brown (1969). Again, entrapment of air during mass movement may produce similar structures, but the extreme (macroscopic) forms figured by Bunting could not have survived lateral movement. Precipitates will be mostly carbonates in the cave situation. Adolphe (1972) has produced concretions experimentally by repeated freeze-thaw. The concretions are characterised by a band of calcite, with an outer fringe of iron hydroxide, a pattern representing the efficient fractionation of the two substances during freezing. Vogt (1977) has recognised such concretions in fossil alluvial deposits formed in cold environments. Cogent structural and contextual arguments are presented to support the climatic interpretation, as well as arguments effectively disproving a pedogenic or primarily illuvial origin. The yellowish white concretions may be up to 2mm thick on the underside of limestone clasts. The phenomenon is particularly common in coarse, rather open deposits and has not been observed in contiguous beds of fine sediment (which are calcareous but not altered or indurated). Bedding is often at a significant angle and the deposits as a whole are permeable, ruling out conditions of waterlogging and gleying. Vogt retains the following features as diagnostic: (1) a zone of decalcification, several tens of microns thick, near the surface of a clast; (2) calcite crystals that penetrate (that are 'rooted') up to c.0.5mm into the clast; (3) banding between calcite (thick) and iron hydroxide (thin), with very good separation of the two. Clasts may be cemented together with friable or crystalline calcite, often containing large quantities of primitive plant detritus (lichens, mosses, algae, etc.), material that decomposes so slowly in a cold environment that

it has been trapped in the growing calcite.

Finally, the disruption caused by minor bodies of segregation ice (very thin lenses and veins) in fine sediments might be expected to produce many fault-like structures and tiny discontinuities in bedding (cf. the photographs in Linell & Kaplar 1966:485).

The term 'convoluted bedding' is used here as a descriptive label, with no genetic implications, to cover undulating and generally irregular bands of sediment. Such structures may represent one, or a combination, of three broad processes:

(1) Distortion of pre-existing bedding. In cold environments, this may occur due to ice segregation (heave and thrust), to the build-up of cryostatic pressure (pore water pressure increase ahead of a freezing front), or to general differential settling of wet sediment during the thaw.

(2) Localised sub-surface translocation of varying amounts of sediment en masse. In a cold environment, this is usually an extension of the causes noted in (1). Cryostatic pressure may inject fluidised sediments into neighbouring layers. During the thaw, differential bulk density of various sediment/water mixes may result in complex interpenetration of beds, or components of beds.

(3) Particle sorting by size and, probably, by shape. In a cold environment, sub-surface frost sorting (infra) may create 'false' bedding, with contorted bands of coarser and finer material.

It may be extremely difficult to separate the various processes in practice, although it may be rather important to try, both for stratigraphic reasons and in order to decide

whether such structures were indeed ice-induced.

Sharp defined periglacial involutions as "aimless deformation, distribution, and interpenetration of beds produced by frost action" (1942:115). This definition would appear to correspond exactly with the concept of 'convoluted bedding' as described above, save, of course, for the environmental connotation. Washburn (1979) argues that fine distinctions in form, with assumed environmental implications (such as whether or not permafrost was present), are very dangerous, and he therefore proposes synonymy between a whole series of terms proposed in the literature, with "periglacial involutions" as the preferred term. The present author fully agrees that an overall term is necessary, but he cannot see why ice-induced structures should be restricted to a periglacial regime. The term 'cryoturbation structures' implies (etymologically) the involvement of ice without the necessity for a periglacial regime. Although 'cryoturbation' was never properly defined (Edelman, Florshütz & Jeswiet 1936), and although different authors have used the term to mean very many different things, 'cryoturbation structures' here refers to the 'aimless deformation, distribution, and interpenetration of beds' thought to be associated with the presence of ice. The qualification 'aimless' is useful, since it allows a (theoretical) distinction between these structures and, on the one hand, more regular features associated with patterned ground and, on the other, similar structures having a preferred orientation (plications, cf. p.345) associated with mass movement. Irrespective of the elaboration of terminology to describe modern occurrences, from known environments, Washburn's warning with respect to fossil occurrences must not be ignored:

> Moreover, contortions appearing very similar if not identical to periglacial involutions can be formed by other processes than frost action, including non-periglacial load casting, slumping (and perhaps other types of mass-wasting), and volume changes whose results can be strikingly like periglacial involutions ... [followed by an extensive list of references]. (1979:172)

'Cryoturbation' has been invoked in literally thousands of reports on cave sediments, with only a handful including even a passing description of the actual structure. For instance, Allsworth-Jones (1978) infers 'cryoturbation' up to at least 20m inside the Szeleta Cave (Hungary) merely on the grounds of edge damage to bones, stone artefacts and limestone clasts. To the present author's mind, all that is indicated by such an observation is some degree of movement, of an unknown cause. Reports of 'frost churning' in deposits that have totally undisturbed lower boundaries would seem to be most dubious. Problems with the concept of 'cryoturbation' will be discussed further in the next chapter. Convincing examples of cryoturbation structures are remarkably difficult to find in the cave context. Despite only cursory description, such structures would seem to be present in the Dzeravá Cave (Czechoslovakia; Prošek 1951) and in the Cueva Morin (Spain; Butzer 1981). In the latter site, Butzer notes that one occurrence may be due to the water retention capacity of a particularly organic-rich layer.

Frost action may sort coarser particles to the surface of sediments. Two hypotheses have been put forward to explain this phenomenon. The 'frost-push' hypothesis, clearly stated by Nansen (1922), relies upon the fact that stones are better thermal conductors than the surrounding sediment and will therefore act as heat sinks once freezing sets in. Ice lenses will form beneath the stones, forcing them up, with infiltration of fines during the thaw preventing complete settling. Beskow

(1930) suggested 'frost-pull'. The descending freezing front catches the top of a stone and the stone is pulled upwards in the general expansion. The cavity below the stone is distorted by increasing pressure and partially filled by fines during the thaw, thus preventing complete resettling. Both 'push' and 'pull' processes probably play some part, and both mechanisms have actually been observed in experiments (cf. Kaplar 1970; Mackay, pers.comm. in Washburn 1979). Elongate particles lying at a slight angle from the horizontal may have their upper end caught by the freezing front and pulled up, creating torque. Such stones may therefore gradually rotate to a more or less vertical position after many freeze-thaw cycles. It is of interest that upward movement of artefacts, in just the same way as rock particles, has been noted by Johnson and Hansen (1974) in open-air contexts. Vertical orientation of stones and artefacts is reported comparatively frequently from caves (cf. Miskovsky 1969), but it is rarely clear whether this was due to frost sorting, mass movement, downward migration through fine sediment, trampling or a number of other processes.

Crack-and-fill structures may originate and develop in various ways. Cracking may be caused by: (1) desiccation in 'swelling' sediments; (2) synaeresis (sub-aqueous shrinkage of originally flocculated colloids and silts); (3) thermal contraction due to rapid temperature drops (only in sediment with a relatively high ice content); (4) heaving of frozen or consolidated sediment. The first two causes are not associated with ground ice, although desiccation may occur in cold climates and may be encouraged by ice segregation. Desiccation cracks may arise in many types of sediment (including matrix-rich gravels), as long as some clay minerals and/or organic matter

are present. Cracks are quite often observed in deeper cave sediments, even when relative humidity is close to 100% and the sediment is still damp to the touch; shrinkage is probably an irreversible process here, reflecting reorganisation of the water/colloid system, akin to synaeresis. The author has never observed clear desiccation cracks in typical (coarse) entrance facies deposits, although major cracks often develop on vertical sections due to progressive failure. Once a crack is open, it may accumulate dry, fine sediment, or (during rewatering or thawing) any type of sediment by slumping. Apart from those formed by irreversible processes, cracks will be seasonal and only open for short periods. Fossil fill structures will therefore tend to be rather narrow, perhaps even microscopic in some fine sediments. Only if the cracking process, once repeated, makes use of the same line of weakness will a major, composite structure develop.

The only larger scale structures known to result from cracking are either found in arid environments (hot or cold), or where ice is at least indirectly involved. Given ground ice, ice-wedges, sand-wedges, soil-wedges or composite ice/sediment--wedges may form. In the fossil state, it may be impossible to decide whether sediment was associated with a wedge whilst it was still active or merely after the ground ice began to degrade. Accordingly, it is safer to refer all fossil structures to the general class 'wedge casts', with the term 'ice-wedge casts' reserved for occurrences where it seems likely that the wedge itself once contained significant quantities of segregation ice.

Wedge casts produced by desiccation or heave cracking are usually quite thin, with relatively fine fill material showing vertical foliation developed as pressure increases

during thawing or rewatering. Observation of modern occurrences (cf. Washburn 1979) suggests that cracking does not recur very many times in the same place and that new cracks open up, eventually leading to a relatively dense pattern of wedge casts. Wedges actually containing ice will only persist for long periods under permafrost conditions. The position of the crack is stabilised by the fact that the lower part of the wedge penetrates the pergelisol and therefore does not degrade seasonally. Casts of this type of wedge may be much broader than those of 'dry' cracks. The fill may be quite coarse (with clasts up to the total width of the cast) and roughly stratified, with slump structures near the boundaries. This reflects infilling during the thaw. Indeed, many casts of this type represent the lower part of the original wedge, which finally becomes an available void only when the permafrost eventually breaks down; the entire fill may thus consist of stratified slump material deposited over a short period. Alternatively, the fill may be quite fine, with complex sub-vertical 'sand veins'. This reflects ice forming in an existing fine fill, any resulting cracks being choked with more sediment before the eventual thaw. There is a more or less continuous gradational series between these two types of cast, well developed examples of the latter type usually being designated as 'sand-wedge casts' as opposed to 'ice-wedge casts' (cf. Black 1976). Foliation may occur in any type of cast in zones with fine sediment. As ground ice warms during the summer, or during the eventual thaw, the sediment mass will expand, so that strata adjacent to a wedge will be deformed with an upward bending of the layers. In cohesive or rigid (clast-supported) sediments, this deformation may persist in the fossil state. In less cohesive sediments, or

in sediments that have low liquid limits, slumping and even multiple faulting later during the thaw may have the reverse effect, with radical downward bending or stepping of the stratification. Stones in peripheral sediments will often be re-aligned parallel to the sides of the wedge, either through slumping or ice-induced processes (supra).

Wedge casts are not always regular in shape. The majority of the sediment filling a fossil structure must be demonstrably derived from above, or from the sediments through which it passes, for the structure to be called a 'wedge cast'. However, it is comparatively common to find casts with irregular extensions, some of which may contain sediment derived from below. These are fluidisation effects (including injections), presumably occurring mostly during the final phase of ground ice decay. Since wedge casts are best preserved if they were formed at a level which was within the original pergelisol, they may only become visible at some depth (the thickness of the original active layer) below the former surface. Care must therefore be taken to refer 'truncated' casts to their correct stratigraphic horizon. When dealing with irregular structures, it should be noted that repeated, faster mass movement in very badly sorted sediment (from gravel to colloids) may cause downward injections, or not obviously upward fluid escape structures, that may sometimes look very like irregular ice-wedge casts (cf. figures in Postma 1983). These structures are usually cylindrical whilst wedge casts are, of course, linear in plan.

Vértes (1959) states that Kerekes found 'ice-wedges' in the Homoródalmáser Cave (Hungary?), but no reference is given, a most frustrating situation since this is the only report of wedge casts in a cave deposit that the present author has ever

come across! Butzer (1981) notes cryptocrystalline calcite in vertical "frost cracks" at the Cueva Morin, but it is not at all clear what is meant by this statement.

Patterned ground results from a combination of any of the processes discussed above, active over a significant area and resulting in a repetition of suites of structures. A single unit in such a pattern may be called a 'cell'. Viewed from above, cells may be roughly linear ('stripes', associated with relatively steep slopes), circular or polygonal. Viewed in cross-section, cells often appear as rough lower semi-circles (a 'festoon' pattern). Poorly developed patterned ground may not have a continuous grid of cells. In the extreme case where conditions for formation are truly localised, a single isolated cell may develop. There are many types of patterned ground (cf. Washburn 1979) but only a few need be discussed here. Because limestone clasts are so common in cave sediments, most cells will be visible because of some kind of sorting. Coarse elements are sorted to the edges of cells and the resulting forms are known as 'sorted circles' or 'sorted polygons', depending upon their plan. The finer material in the centre may be totally surrounded (underneath as well) by coarser material, the latter usually oriented at right angles to a radius from the centre of the hemisphere. Sometimes central finer material has been injected from below and the orientation of coarse material may be affected near the injection point. If the entire feature is predominantly composed of coarse elements, oriented at right angles to any radius from the centre, it may be called a 'stone pit' (often with a large central stone) or a 'stone nest'. The term 'garland' should not be used to describe a stone nest in section (<u>contra</u> Laville 1975;68), since

this term properly refers (in all European languages) to stone-bordered solifluction lobes as seen in plan. Roughly hemispherical patches of finer sediment may occur, even on steep slopes, in otherwise coarse and open deposits; these are called 'debris islands', an equivocal term to the present author's mind. 'Debris islands' appear to be related to frost action in some cases, but the exact mechanisms of formation are unclear. Any of the above types of cell may also have wedge casts at their peripheries (between cells, in a continuous pattern), as long as there is some finer sediment in which they could form. Most 'closed' cells (circles and polygons) tend to become more elongated at steeper slope angles, eventually losing cross-slope structures to become stripes. All types of patterned ground that do not include classic ice-wedge or sand-wedge casts may occur well outside permafrost areas (cf. Washburn 1979).

Patterned ground, usually consisting of stone nests, has been reported from a few caves. Simply from textural constraints, it is possible that such structures are the most likely forms of frost-associated phenomena in most cave sediments. A most impressive example of contiguous cells, with some fine material in the centres, is illustrated (in photographs and a drawing) by Alimen (1950) from the shelter of Mouthiers (Charente; cf. also Alimen & David 1949). That closed cells are present is shown by structures visible on two sections cut at right angles, although the claim of polygons (rather than circles) is not supported. Clasts, averaging c.10cm, are vertical between cells and swing round smoothly in 'festoons' under each cell. Cells are about 140cm across and 80cm deep. Most cells have larger blocks (40-80cm) at odd angles near their centres and this extra loading may have helped to form the

pattern. Similar patterned ground was described by Schandler (1931) from the Eislueg Cave (Austria). At the Pech-de-l'Azé II (Dordogne), Laville (1975; cf. especially Plate 30, p.68) reports contiguous stone nests, this time with interstitial material restricted to clayey coatings on limestone clasts. Cells are up to 100cm wide and 40cm deep, with the effects of deformation showing slightly even in the underlying deposits. The nests are rather more variable in dimensions and stone orientation than at Mouthiers, but there can be no doubt as to their origin. At the shelter of Le Moustier (Dordogne - not to be confused with Mouthiers), Laville (ibid.) notes that Layer I consists of a badly sorted and massive sediment, despite the fact that the mineral suite could only have been derived from the River Vézère. Limestone clasts are rounded, but not 'porous' (not corroded), so that Laville interprets the deposit as being due to 'cryoturbation'. Alimen and David (1949) appear to have been referring to this deposit when they recorded the sighting of a stone nest, 70cm deep and 50cm wide. Laville does not mention their observation. An isolated stone nest, with a finer centre, has also been identified at Pixie's Hole (Devon; cf. p.852).

Although reports are still all too rare, stone nests, whether isolated or in groups, may have something in common in the cave context. First, they normally occur where drainage is comparatively restricted, either by the proximity of underlying bedrock or by underlying relatively clayey deposits. Second, they tend to occur, or to be best developed, under rock overhangs or below roof fissures, locations which might receive water right up until the stage when freezing was quite well advanced. Vertical piping alone cannot possibly be the cause of even isolated stone nests; examples of suffosion pipes (cf. p.446)

observed by the author never have consistent macrofabric, they are always much deeper than they are wide, they often cross otherwise undisturbed strata and their overall shape is usually highly irregular.

An exhaustive search of the European literature on cave sediments would probably only produce two or three dozen instances of certainly ice-induced structures. It has been argued that this is mainly due to the special circumstances in cave deposits. Nevertheless, although identification of ice-induced structures will nearly always be difficult in the cave situation, it appears to the author that more information could be collected through more careful observation. This proposition calls for a state of mind whereby the action of ground ice is no longer just taken for granted. The best proofs rest upon the identification of a whole series of structures, at different scales, observed over the widest possible exposures of a single bed or group of beds.

8.3. Rock Shattering

The growth of ice in more restricted spaces may cause the failure of any surrounding rigid material. This process has been called 'frost wedging' (an equivocal term, despite the fact that it is the most commonly used amongst geomorphologists), 'gelifraction', 'congelifraction', 'frost riving', 'frost splitting' and 'frost shattering'. The adjective 'cryoclastic' has often been used in the cave situation, and the word has analogues in most European languages. The author has only seen the substantive forms cryoclastie and cryoclase in French. Since even the term 'clastic' is a neologism, the forms

'cryoclastis' (cf. "thermoclastis") and 'cryoclasis' (cf. "cataclasis") would seem to be equally acceptable in English. The terms 'thermoclastic' and 'thermoclast(ic)ism' (cf. Texier 1973; Laville 1976a; Campbell 1977; Laville et al. 1980) are not acceptable, since there can be confusion with insolation weathering (thermoclastis sensu stricto).

There is a massive literature on experimental cryoclastis, much of it concerning limestone (cf. Washburn 1979 for a detailed bibliography). Through such experiments, the major controls have been identified, at least with respect to shattering of pre-existing clasts (rather than bedrock). Adequate water is a prerequisite for cryoclastis and the water must be able to penetrate the rock. This may occur along discontinuities (bedding planes, joints, fractures, primary or metamorphic planes of fissility, etc.) or through the pore network. The grain size of sound rock controls the pore size, the thermal conductivity and also, to some extent, the strength of the rock. These factors, together with the temperature and rate of temperature change, affect the way in which water freezes within the rock. Finer pores will only allow freezing at lower temperatures; water may be expelled to the surface if the latter is not frozen (inhibiting shattering) or conducted to centres of ice segregation if the surface is frozen (promoting shattering). Very fast temperature decline may block off pores even in coarse grained rock, thus promoting shattering. On the other hand, slow temperature decline may allow fine grained rocks to take in even more water from their surroundings to be used to build segregated crystals, thus also promoting shattering. By varying the temperature regime (amplitude, frequency and rate of change), the rock type (grain size and porosity, strength,

etc.) and the water availability (water baths covering different proportions of the rock sample, prior saturation of the sample with no external water during freezing, etc.), experimenters have produced results suggesting that cryoclastis is an extremely complex process. The question is to what extent these findings can be usefully applied to the interpretation of fossil limestone scree.

Natural cryoclastis has often been divided into two classes: primary cryoclastis is the shattering of bedrock at a free (air) surface and secondary cryoclastis is the shattering of limestone clasts once they have fallen and become part of the sedimentary fill. This division may be justified as follows: (1) water availability will be radically different in the two situations; (2) the temperature regime will be radically different in the two situations; (3) the dominant route by which water penetrates (planar discontinuities or pores) will be different in the two situations; (4) primary cryoclastis involves cooling more or less from one direction, whilst secondary cryoclastis may involve cooling from large areas of the surface of a clast. Given the recognition of the critical factors by the experimenters, we might expect primary and secondary cryoclastis to have different controls and different products. There are intermediate situations between the two classes (shattering of bedrock buried under sediment, shattering of rockfaces just above a moist sediment contact, etc.) which require explicit description. In fossil deposits, material may be present that has suffered primary then secondary cryoclastis, that has suffered only primary cryoclastis, that has suffered only secondary cryoclastis (the clasts having been originally formed by processes other than cryoclastis) or that has not suffered any form of cryoclastis.

Our ability to disentangle the 'history' of a clast will depend upon whether or not clasts produced by the different types of cryoclastis are significantly different in some way and whether they are also different from clasts produced by other processes. Just some of the processes which have been suggested as causing, or contributing to, rock shattering include: earthquakes (cf. Escalon de Fonton 1971); preferential solution along structural planes (cf. Lais 1941; Cogley 1972); bacterial attack along mineral veins (cf. Ollier 1969); crystal wedging by calcite and other salts (cf. Ollier 1969; Evans 1970; Goudie 1974; McGreevy 1982); rupture of unsupported rock masses and rupture when support (water, sediment) is rapidly removed (cf. Renault 1968; Carson & Kirkby 1972); residual stress release (cf. Gerber & Scheidegger 1969; Carson & Kirkby 1972); root wedging (cf. Sparks 1972); hydration (cf. Dunn & Hudec 1972; S.E. White 1976b; Wilhelmy 1977); thermal or insolation weathering (cf. Ollier 1969); fire (Birkland 1974); slaking, involving the build-up of oriented water molecules during wetting and drying (cf. Ollier 1969); loading by glaciers (cf. Ford 1976); rockfall impact (cf. p.319); cataclasis or crush-brecciation (cf. p.608); hydrostatic pressure (infra). Together with cryoclastis, these processes may act in various combinations to produce shattering, with differential effects depending upon often extremely localised properties of the rock (cf. Frank 1977). In order to assess the various probabilities, we will need considerable information on the limestone (porosity, structure, strength, thermal properties, mineral veins, colloid content and disposition, etc.), on the morphology of the site (geometric relationships between limestone structure and changing outcrop form, etc.), on various properties of the deposits (water retention capacity,

frost-susceptibility, chemistry, etc.), on the prevailing general climate (temperature regime, precipitation, cloudiness, etc.) and on site specific parameters (aspect, altitude, proximity to the sea, etc.). Finally, we must be sure that a clast was formed in the environment which we have inferred and was not transported, at some unknown time after its formation, to its observed position. To the present author's mind, all this seems to be asking rather a lot of a sample of clasts.

The attitude taken in the following discussion is not that most limestone scree is unrelated to cryoclastis, or to a combination of cryoclastis and other processes. On the contrary, the general empirical link between increasing scree production and increasing number of frost cycles is adequately proven. What worries the present author is the tendency for slight changes in the proportions of scree, and in the size and shape of clasts, to be interpreted in direct climatic terms. On p.1040, the Sandy Cave Earth at Bacon Hole (Gower) will be described in detail. This is a roughly tabular deposit, about half a metre thick, with a lateral extent of at least 30m and lying at a very gentle angle. Some sandy matrix is present but the deposit is totally dominated by loose, angular scree, from grit to clasts of 20cm diameter and more. Yet this deposit contains a classic ipswichian fauna, without the slightest possibility of derivation. The automatic inference of cryoclastis from the mere presence of limestone scree must be abandoned. Furthermore, even if cryoclastis seems a likely origin, the automatic assumption that the only significant controls are climatic must also be abandoned.

Before continuing with an examination of limestone shattering, one situation in which cryoclastis is unlikely to

be involved should be noted, that is, the shattering of extensive stalagmitic floors. With reference to Joint Mitnor Cave (Devon; cf. p.789), Sutcliffe stated:

> The water-laid sediments ... and the talus deposits ... were locally overlain by a stalagmitic floor, which reaches a thickness of about eighteen inches [c.45cm] in the western extension of the cave. This floor had been extensively shattered by some process (plate 1b [a crystalline floor shattered into huge slabs in the nearby Reed's Cave]) and had been partly recemented by further deposition. Near the [artificial] entrance several large blocks of limestone (the largest ten feet [c.3m] long) had fallen from the roof of the cave onto the talus deposits. The cause of the shattering of the stalagmite and the fall of the blocks is uncertain, but it could be interpreted as the result of a period of frost shattering, more recent than the accumulation of the interglacial talus deposit. (1960:17)

Sutcliffe later revised this opinion:

> Three possible causes are worthy of consideration:-
> 1. Settling of the cave [deposits] following the fall of the water table.
> 2. Frost heaving during the Last Glaciation.
> 3. An earthquake.
> This fracturing extends throughout the entire cave system and is far better developed in Reed's Cave, where the tips of many of the stalactites have been broken and partly renewed. In view of this the third explanation seems the most feasible. (1966:3)

However, Curry (1977) has returned to the original frost shattering hypothesis.

The present author is of the opinion that Sutcliffe was nearer the truth when he mentioned the water table. Crystalline stalagmite of this sort is immensely strong when it occupies the entire width of a cave passage. Settling of underlying deposits will often leave a floor intact, even when there is much additional sediment above the stalagmite, which, in this case, there is not. If cryoclastis (or an earthquake) had been the cause of the stalagmite shattering in the deep Joint Mitnor--Reed's Cave system, one would expect extensive rockfall material above the floor, material which is not in fact present. As at other sites (eg. Rhinoceros Hole, p.975), this shattering is almost certainly due to the build-up of hydrostatic pressure in the sediments underlying the floor. The very nature of caves

means that, once the drainage is blocked by stalagmite (or any other impervious sediment), eventual disruption by hydrostatic pressure is almost inevitable (cf. Hanwell & Newson 1970; Warwick 1976). When shattering is observed throughout an extensive system, and there are no signs of a major flood event, there is no reason to suggest that the shattering must itself represent a single event. The spatial association of a few roof blocks does not necessarily imply a common cause.

Limestone clasts are necessarily a subject of great interest to all cave sedimentologists. Laville, basing his arguments mainly on the work of Tricart (1965) and Malaurie (1968), discusses cryoclastis in the following terms:

> Une forte teneur du sédiment en éboulis peut être attribuée à une phase de gélivation, mais il convient d'en déterminer les modalités et l'intensité. (Laville 1975:30)
>
> [Trans.: A high proportion of scree in the sediment may be attributed to a phase of frost shattering, but it is advisable to determine the conditions [style] and intensity.]
>
> Deux cas limites peuvent être envisagées.
> - Considérons une période froide, caractérisée par un cycle de gel de type annuel, intense et prolongé. Le résultat sera une gélivation du toit de l'abri en larges fragments; il y aura sédimentation concomitante de ces éboulis, sédimentation qui conduira à une protection d'autant plus rapide et efficace qu'elle sera alimentée par un gel plus intense. La gélivation secondaire des éboulis sera d'autre part limitée par une réduction de leur teneur en eau: par suite de l'évacuation immédiate des eaux de fonte entre les vides interstitiels, du fait que l'ascension par capillarité des eaux pouvant imprégner les couches profondes du remplissage sera genée par la présence de ces vides interstitiels.
> - Considérons maintenant le cas d'une ablation peu importante, portant sur le même surplomb rocheux, donc sur un matériau de même texture et de même structure que précédemment, ablation limitée par un gel modéré, de type journalier: il y aura libération de petits éléments et émiettement de la paroi rocheuse. La sédimentation étant peu active, les éboulis seront plus directement soumis aux agents extérieurs (en particulier le gel); ils se trouveront mêlées à une forte proportion de sables, de limons et d'argiles dont la présence contribuera à maintenir l'humidification des débris calcaires: autant de facteurs favorables à une nouvelle fragmentation cryoclastique, limitée certes, mais dont le résultat sera une nouvelle réduction granulométrique des débris. (ibid., p.25)
>
> [Trans.: Two limiting cases may be envisaged.
> - Consider a cold period, characterised by a freezing cycle of the annual type, intense and prolonged. The result will be frost shattering

of the roof of the shelter, giving large fragments. There will be accompanying sedimentation of this scree, sedimentation which will lead to a protection [cover] all the more rapidly and effectively as it is fed by more intense freezing. Secondary cryoclastis of the scree, on the other hand, will be limited by a reduction in its water content, due to the immediate evacuation of melt water through the interstitial voids and to the fact that any water impregnating the deep layers of the fill will be prevented from rising by capillarity because of the presence of these interstitial voids.
- Consider now the case of weak breakdown, affecting the same rock overhang, thus a material of the same texture and structure as before, breakdown limited by a moderate frost of the diurnal type. There will be production of small elements and crumbling of the rock wall. Since sedimentation will not be very active [rapid], the scree will be more directly subjected to external agents (frost action, in particular). The scree will be mixed with a large proportion of sand, silt and clay, the presence of which will contribute to the maintenance of humidity around the limestone debris. All these factors will favour a further cryoclastic breakdown, limited certainly, but resulting in a further grain size reduction in the debris.]

This self-contained model is, at first sight, extremely attractive. The calibre of the scree will be directly proportional to the degree and persistence of winter cold. There are no factors which will introduce a random element and no factors which will reverse the proportionality. However, the model is based upon a number of assumptions which are rather important when comparing the layers of a real sequence:

(1) Bedrock is totally homogeneous.

(2) All site-specific factors (such as received sunlight, geometry of the shelter, etc.) are constant.

(3) No material, either scree or finer sediment, is ever removed from, or added to, the deposit other than by cryoclastis.

(4) The rate of primary production of large clasts is automatically fast enough to ensure that coarse deposits will always build up faster than finer deposits.

(5) There is never any interference with the system from snow or persistent ground ice.

(6) The moisture regime is irrelevant for primary

cryoclastis, only temperature being of importance.

(7) Environmental availability of water is irrelevant for secondary cryoclastis, only the sediment's retention capacity being of importance.

(8) Freezing, and thus the possibility of secondary cryoclastis, never occurs significantly below the current ground surface.

There are very few situations in which the present author would be willing to accept any of these assumptions. Even the basic proposition that stronger frosts produce larger primary debris is unproven.

Farrand, working at the Abri Pataud (Dordogne), has tried another approach, based mainly upon the experimental data made available at the Centre de Géomorphologie, Caen (cf. Guillien & Lautridou 1970).

> In summary, these experiments on frost weathering of limestone show that smaller fragments will be produced by:
> i. more freeze-thaw cycles;
> ii. greater intensity of freezing; and
> iii. greater degree of [water] saturation of the rock.
> And, a greater quantity of debris per freeze-thaw cycle will, in general, be produced either by:
> i. greater intensity of frost, or
> ii. greater degree of saturation of the rock.
> Thus, the size and abundance of frost debris is a complicated function of time (number of freeze-thaw cycles), temperature (intensity of frost) and moisture (degree of saturation), with the overall result that, for any given limestone, the size of the resulting debris is inversely proportional and the quantity of debris directly proportional to each of these factors. Unfortunately, it is impossible to evaluate the effect of each of these factors separately in the absence of independent information about one or more of them.
> One limitation in the application of the experimental results of [to?] rock-shelter sediments concerns the upper size limit of frost-derived fragments. The experiments were run on solid blocks of limestone about 100mm. on a side, thus imposing an artificial limit on the largest fragments removed by freezing. On the limestone wall of a rock-shelter, however, the only limit is that imposed by the spacing of cracks and bedding planes, and these cracks may be much more widely spaced than 100mm. To add to this complication, it must be emphasised that two different kinds of fragments are produced by the freezing and thawing of a natural bedrock outcrop. The first of these may be called "frost spalls"; they are relatively small chips of rock removed from a solid rock surface by the growth of interstitial ice. The experiments cited above deal

exclusively with such fragments, and their longest dimension is commonly less than 30mm. However, water may freeze within any natural crack or system of cracks within the rock, and the entire block of rock bounded by these cracks may thus be detached. We can call these fragments "frost blocks". Frost blocks are not dealt with in the experiments, and the largest size of the blocks that may be detached in this fashion is not really known. They are obviously limited by the natural spacing of the cracks, but it is intuitively questionable whether freezing of water in cracks could detach blocks as large as one meter cube, such as are commonly seen in rock-shelter deposits in the Dordogne. On the other hand, it might be considered reasonable to suggest that blocks 10, 20, or 30cm. on a side might be detached by freeze-thaw activity.

[...] [It can be seen from the experiments that] the initial frost spalls themselves, in turn, are broken down into smaller fragments by the ensuing freeze-thaw cycles. This process does not continue indefinitely, however; the ultimate particle size distribution tends to stabilize after a certain number of cycles, being controlled by the size of the original components of the limestone, such as sand grains, fossils, ooliths, muddy matrix, etc. This stabilized particle-size distribution is attained relatively rapidly for many limestones that were studied, i.e., after only some 100 to 200 freeze-thaw cycles. [...]

A crude approximation of freezing conditions in the Abri Pataud is based on the assumption that during Last Glacial times there were about 100 freeze-thaw alternations per year at the ground surface, but only one cycle per year at the depth of greatest penetration of frost into the ground. This depth remains unknown, but it is reasonable to assume that it was equal to or greater than 100mm. Then, let us assume that the uppermost 100mm of the accumulating éboulis [scree] in the Abri Pataud underwent on the average 50 freeze-thaw cycles per year. Given a sedimentation rate of 1 to 2mm. per year [a figure calculated from C14 dates], an initial layer of éboulis 100mm. thick would have been subject to freezing conditions on the average of 50 times per year for some 50 to 100 years, making a total of 2500 to 5000 cycles. The experiments reported by the Caen laboratory ... span only 200 freeze-thaw cycles at the most. Since the Abri Pataud sediments were exposed to from 10 to 25 times as many cycles, it seems reasonable to make comparisons with the experimental results.

[...] In a general way one could conclude that an éboulis horizon with smaller fragments is an indication of a greater intensity of frost activity, although one cannot determine whether there were more freeze-thaw cycles per unit time or a uniform number of freeze-thaw cycles but of variable amplitude.

[...] [The] size distributions of the great majority of the Abri Pataud éboulis match quite well the size distributions of the experimental frost debris, especially that of calcarenites apparently similar to the Abri Pataud bedrock. [...] At least two modes are evident in nearly all the histograms, and in many of them three distinct modes can be seen, ... These modes fall into the following size classes: 64 to 16mm., 8 to 4mm., and 0.250 to 0.125mm. The smallest of these modes ... corresponds to the principal size of the sandy particles constituting the local bedrock .... The two larger modes both are composed of fragments of the local bedrock, and both fall into the range of particle sizes obtained in the freeze-thaw experiments mentioned above.

[...]

In the Abri Pataud samples, ... , those samples with a mode near 100mm [an apparent mode, since 100mm is the largest diameter considered in the formal size analysis] are all found in beds of éboulis secs (openwork rubble), ... Moreover, a number of other samples with a prominent mode between 64 and 32mm. also come from beds of éboulis secs ....

> Furthermore, most of these éboulis secs are found in association with prominent rockfalls which are represented by huge blocks of limestone generally one to several meters in length. [...]
> Therefore, it appears that the Abri Pataud éboulis can be explained as follows: (i) the coarsest mode, >16mm., appears to be primarily the result of shattering of roof-blocks as they collapse, although there probably is some contribution to this mode by freeze-thaw activity; (ii) the two finer modes coincide exactly with the granulometric results of freeze-thaw experiments cited previously; and (iii) the finest mode, 0.250 to 0.125mm., is mainly the result of the reduction of the bedrock to its ultimate components by freeze-thaw activity, but it has been augmented to a greater or lesser degree by a windblown contribution from the floodplain, as discussed [elsewhere in the report].
> (Farrand 1975:53-4)

Farrand's approach is very much more detailed than that of Laville. Certain aspects of this work are worthy of note. The approach is site-specific, which allows both a comparison with exact experimental data, derived from study of a similar rock type, and a modification of general principles in accordance with other features of the deposits. Problems with material coarser than the experimental samples are acknowledged and the need to discover whether finer debris (calcite sand, in this case) is mixed with material from other sources is recognised. Perhaps the most impressive feature of Farrand's work is his refusal to give undue significance to numerical results. The final reconstruction of the supposed effects of cryoclastis is based upon the presence of prominent modes and not upon minor, and probably random, differences in percentages or summary statistics.

However, there are still major problems of which, no doubt, Farrand was aware. As he himself says, "it is impossible to evaluate the effect of each of these factors [time, temperature and moisture] separately in the absence of independent information about one or more of them."(supra). Farrand makes by far the most concerted effort to understand the textural role of cryoclastis in a real cave situation that the present author

has come across but, ultimately, the degree to which he has succeeded is questionable. Despite the fact that Laville's approach is much too generalised, he recognises, at least in theory, the importance of the fine matrix and the differential sedimentation rate to be expected with different calibre scree. Farrand does not integrate these factors into his model. His argument concerning sedimentation rate is designed only to support his contention that most deposits will have been subject to enough frost cycles for an equilibrium with the prevailing climate to have been achieved and, thus, that the experimental data are applicable to this site. But, coarse scree may still be the result of the absence of an adequate water supply rather than the paucity of frost cycles. Furthermore, the choice of 10cm as the likely depth of frost penetration would seem to be arbitrary and something of an underestimation (cf. section 8.2.), even though this is claimed as a minimum. For instance, samples AP 6-5 and 6-7 were taken only c.25cm vertically apart, with sample AP 6-6 between them. These samples are said to show abundant freeze-thaw (6-5, c.40% fine matrix <2mm), rockfall alone (6-6, c.15% matrix), and rockfall and particularly strong freeze-thaw (6-7, c.30% matrix), based upon the presence or absence of the two coarser modes. Is it really valid to expect sub-surface cryoclastis not to interfere with such fine stratigraphic divisions? Is it impossible that the shattering in AP 6-7 (which is underlain by slightly finer sediments, AP 6-8) might have occurred, at least in part, during the same period as that in AP 6-5, with the intervening AP 6-6 suffering little modification because of lack of a humidifying matrix? The present author believes that Farrand has gone about as far as is practical with particle size data alone but, as in most

situations, a textural parameter will not provide unequivocal solutions. This would also appear to be the opinion of Guillien, an internationally acknowledged expert on the subject, who is here writing in an archaeological journal in order to try his best to communicate:

> Les fractions grossières de ces remplissages, les horizons grossiers de ces remplissages ont été présentés souvent comme témoignant d'épisodes climatiques d'autant plus sévères que les éléments de ces fractions ou de ces horizons sont plus grossiers: il suffirait, avance-t-on souvent, de mesurer la plus grande dimension de chacun des éléments compris entre 10 et 1 cm, de la mesurer au cm près, pour obtenir quelque chose comme une échelle de température à divers moments de l'histoire de l'édifice détritique. Or il faut, tout au contraire, affirmer que ces fractions grossières n'attestent clairement qu'un rapport de simultanéité entre un état de la température de l'air et un état de l'humidité de la paroi: la température de l'air et celle de la surface rocheuse oscillaient autour de 0°C alors que la teneur en eau de la paroi dépassait la teneur en eau critique .... On définit par là une saison critique, peut-être tout à fait brève, mais qui ne présentait pas nécessairement des températures sensiblement inférieures à 0°. A certains jours de cette saison critique, à d'autre périodes de l'année que cette saison critique, la température de l'air peut-être a été très basse: mais le fait est à établir, et il ne peut être établi par le simple classement dimensionnel de fragments calcaires dont les dimensions sont comprises entre 10cm et 1cm. A elle seule la présence d'une fraction grossière importante ne démontre que la paralysie ordinaire des mécanismes de réduction, que l'incapacité où la cryoclase a été d'exploiter les passages de la température par 0°C. (1970:232)
>
> [Trans.: The coarse fractions of these fills, the coarse layers of these fills, have often been presented as testimony to climatic episodes of a severity increasing in proportion to the coarseness of the elements of these fractions or of these layers. It is often suggested that it would suffice to measure the greatest dimension of each element between 10 and 1cm - to measure it to the nearest centimetre - in order to obtain something like a scale of temperature representing various moments during the history of a detrital formation. Now, it must be stated that, quite to the contrary, these coarse fractions only testify clearly to a relationship of simultaneity between a temperature state of the air and a moisture state of the rockwall: the air temperature and that of the rock surface oscillated about zero whilst the water content of the rockwall was greater than the critical water content .... This defines a critical season, perhaps quite brief, but which did not necessarily entail temperatures markedly below zero. On certain days during this critical season, and during periods of the year other than this critical season, the air temperature was perhaps very low. However, this must be demonstrated and it cannot be demonstrated by the simple dimensional classing of limestone fragments between 10 and 1cm. On its own, the presence of an important coarse fraction shows only the unremarkable paralysis of the reduction [comminution] mechanisms - only the incapacity of contemporaneous cryoclastis to exploit temperature oscillations about zero.]

This statement is just as valid for secondary cryoclastis

as it is for primary shattering. Although it is designed to refute the sort of approach advocated by Laville, the statement also shows up some of the (unavoidable) weaknesses in Farrand's arguments.

Farrand knew that cryoclastis ultimately produces fine particles and he attempted as best he could to take this fact into account. Most experiments reported in the literature resulted in the production of large quantities of sand, silt and even colloids, depending upon the lithology, in addition to larger clasts. The difficulty with finer clasts is that they are difficult to identify and they are both mechanically and chemically unstable. Except in the situation where constituent particles are intrinsically rounded, particles of sand size produced by cryoclastis should be angular. Only by microscopic examination can such small clasts (as opposed to other carbonate grains) be recognised. The presence of such clasts is proof of some type of shattering (not necessarily cryoclastis) but very small amounts of movement or alteration will suffice to obscure the clastic origin. It is interesting that Goldberg(1979a), counting sand sized clasts in thin sections from the deposits of the Pech-de-l'Azé II (Dordogne), recognised most maxima in apparently cold and dry contexts. One wonders whether this impression of 'driness' was enhanced by weakness, or lack, of post-depositional modification. Note that there is no clear relationship (either directly or inversely proportional) between Goldberg's counts of fine clasts and the figures given by Laville (1975:76) for either the coarse component or for 'plaquettes' (supposedly frost-related forms, infra) in the same sediments, a fact which recalls a statement by Guillien:

Finalement le rapport en poids des fractions fine, moyenne et grossière,

> reste sans signification pour autant qu'à l'analyse granulométrique ne sont pas associés une étude physico-chimique du sédiment et un essai au gel de la roche-mère. (1970:232)
>
> [Trans.: Finally, the weight proportions of the fine [silt and colloids], medium [sand] and coarse [gravel] fractions will remain without significance as long as the granulometric analysis is not accompanied by a physico-chemical study of the sediment and testing of the bedrock's response to frost.]

Thus, we are once again faced with the problem that, although we know that large quantities of fine clastic sediment will certainly be produced by frost action, there is no obvious way of proving the point in most real situations.

Several authors have tried to identify different types of cryoclastis through the shape of limestone clasts. Laville, alluding to the two end members of the range of cryoclastis recognised by him (the 'annual' and 'diurnal' frost cycles, supra), states:

> Nous avons en effet distingué, parmi les cailloux calcaires, deux formes de fragmentation correspondant aux deux cas extrêmes que nous venons de décrire.
> - Les cailloux gélivés, caractérisés par un éclatement total en deux ou trois larges fragments. Ces formes sont fréquentes parmi les éboulis de fortes dimensions; elles sont plus rares dans les dépôts de granulométrie plus fine. Une telle fragmentation suppose un gel intense, fragmentation cependant rapidement freinée par les caractéristiques mêmes du gel qui l'a favorisée.
> - Les cailloux fissurés, caractérisés par un réseau superficiel de fines craquelures. Nous avons identifié de telles formes dans des sédiments de granulométrie fine, à matrice sablo-argileuse abondante. L'ensemble des caractéristiques sédimentologiques des dépôts auxquels ces cailloux fissurés sont associés nous a conduit à les interpréter comme le produit de phénomènes de gel et de dégel de faible intensité et de faible amplitude, favorisés d'autre part par la présence d'une matrice fine abondante. (1975:25)
>
> [Trans.: Indeed, we have distinguished two forms of fragmentation, corresponding to the two limiting cases which we have just described, among the limestone clasts.
> - Frost-shattered stones, characterised by complete splitting into two or three major fragments. These forms are frequent among scree of large dimensions; they are rarer in deposits with finer particle size distributions. Such fragmentation implies intense freezing, fragmentation which is nevertheless rapidly inhibited by those same characteristics of the frost which [first] favoured it.
> - Fissured stones, characterised by a superficial network of fine cracks. We have identified such forms in sediments of fine particle size distribution, with abundant sand-clay matrices. The total sedimentological characteristics of the deposits with which these

fissured stones are associated have led us to interpret them as the product of freeze-thaw phenomena of weak intensity and amplitude, favoured moreover by the presence of an abundant fine matrix.]

Laville goes on to describe how these two types are recognised in practice:

Recherche des cailloux gélivés.
[...]
Ces fragments ne demeurent jointifs que lorsque le dépôt auquel ils appartiennent n'a fait l'objet d'aucun remaniement postérieur. Ils apparaissent généralement légèrement décalés les uns par rapport aux autres et sont parfois rendus solidaires par la cimentation secondaire de calcite diagénétique.
Nous avons classé dans la catégorie des cailloux gélivés des éclats de roche polyédriques dont l'une des faces, à arêtes plus tranchantes, a conservé l'aspect d'une cassure: il s'agit de fragments de cailloux gélivés, dispersés par les remaniements postérieurs ou dissociés lors du prélèvement des échantillons.
Recherche des cailloux fissurés.
[...]
Il s'agit généralement d'éléments de petite dimension, fréquemment émoussés, dont toute la surface est parcourue par un réseau de fines craquelures plus ou moins béantes. Ils présentent l'aspect d'une mosaïque de fragments anguleux, séparés par des fissures de 1 à 3 mm. de profondeur .... (ibid., p.32)

[Trans.: Study of frost-shattered stones.
[...]
These fragments remain together only when the deposits to which they belong have been subject to no later disturbance. In general, any fragment appears slightly displaced with respect to others and they are sometimes joined together by secondary cementation by diagenetic calcite.
In the category of frost-shattered stones we have grouped polyhedral clasts that have one surface, with sharper edges, that has retained the aspect of a break; these are fragments of frost-shattered stones, dispersed by later disturbance, or dissociated during sampling.
Study of fissured stones.
[...]
These are generally elements of small dimensions, often edge-rounded, and with their entire surface covered in a network of fine cracks, which have opened up [which are 'gaping'] to a greater or lesser degree. They appear as a mosaic of angular fragments, separated by fissures of 1 to 3mm depth ....]

Laville recognises one further diagnostic form:

Recherche des plaquettes de gel.
Dans certaines conditions de température, d'humidité, d'exposition et de lithologie, la gélivation de la roche en place conduit à la formation de plaquettes.
Ces plaquettes telles qu'elles apparaissent avant tout remaniement, présentent un aspect typique, avec des arêtes vives et une face plus fraîche correspondant à la face d'éclatement ....
Les cailloux correspondant à cette définition sont extrêmement rares dans les remplissages d'abris par suite des modifications physiques et chimiques qu'ils ont subies après leur formation. C'est pourquoi nous avons assimilé à cette catégorie des formes moins nettes, parfois

émoussées, mais cependant identifiables par leur aplatissement. Nous avons donc classé sous la rubrique des plaquettes tout caillou présentant une épaisseur inférieure au quart de sa plus grande largeur, sans tenir compte de son aspect. (ibid., pp.31-32)

[Trans.: Study of frost 'plaquettes'.
Under certain condition of temperature, humidity, aspect and lithology, cryoclastis of in situ rock leads to the formation of 'plaquettes'. Before any disturbance, these 'plaquettes' present a typical appearance, with sharp edges and one fresher surface corresponding to the surface of fracture ....
Stones corresponding to this definition are extremely rare in the fill of rockshelters because of physical and chemical modification which they have suffered after their formation. This is why we have assimilated into this category less clear forms, sometimes edge-rounded, but nevertheless identifiable by their flatness. We have therefore classed under the heading of 'plaquettes' any stone having a thickness less than a quarter of its greatest width, irrespective of its appearance.]

Looking through Laville's descriptions of individual sites, we find that 'plaquettes' occur at nearly all of them, but that 'frost-shattered' stones and 'fissured' stones appear at only a few. Curiously, the latter two types are never reported to be present in the same layer; indeed, they are never present in the same major sedimentary unit.

In order to examine Laville's suggestions concerning the matrix associated with 'frost-shattered' and 'fissured' stones, the percentages of these types, as observed in 93 samples from Combe-Grenal (Dordogne; ibid., Plate 72, p.149), have been plotted against percentages of fine matrix (<2mm) in fig. 22 of the present text. 'Plaquettes' are, of course, said to be primary clasts, which should not be affected by the matrix; their percentages have therefore been plotted in fig. 22 as a control. Since the numerical data are derived from Laville's diagrams, accuracy in fig. 22 is only to the nearest 5%. The present author is unable to see any significant patterning in fig. 22 : at this site at least, Laville's suggestions concerning the matrix are not supported. Indeed, if anything, 'fissured' stones seem to present the reverse relationship with the matrix

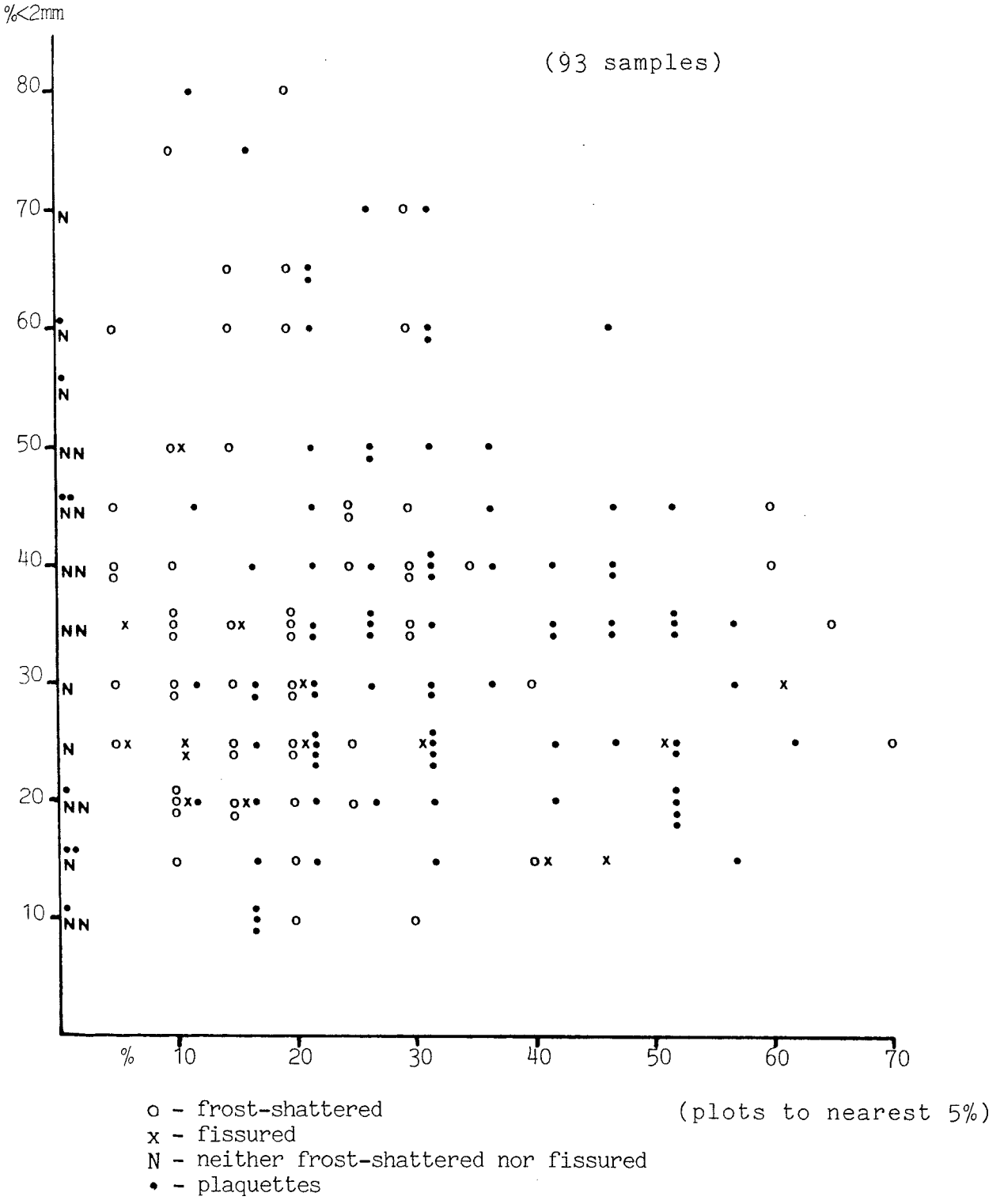


Fig.22 Condition of Limestone Clasts - Combe-Grenal

(data from Laville 1975)

to that suggested by Laville. Since the matrix is the linchpin of Laville's climatic argument, some doubt as to the origin of these types is permissible.

A closer look at these three supposedly frost-related types is necessary.

'Frost-shattered' stones, by Laville's practical definition, are merely compact, polyhedral forms, showing one fresher face. By Laville's own admission, the important structural relationship between fragments, which if observed in situ would at least prove that fracturing had occurred within the sediment (though not necessarily due to cryoclastis), is destroyed during sampling. No explicit reference is made to observed contiguous fragments during discussion of individual sites by Laville. In that a fresher surface will not be present on clasts that have later been chemically altered, the prediction that 'frost-shattered' stones will only have formed under cold conditions is self-fulfilling. In any case, the present author has yet to see a scree body in which the majority of the upper surfaces of clasts are not 'fresher' than the lower surfaces. Laville's insistence on the use of textural data in this case, highlights his lack of understanding of the importance of structure.

'Fissured' stones are said to be due to weak but frequent frost cycles. Apparently identical forms have also been recognised by Miskovsky (1972, Plate 1, p.111; compare with Laville 1975, Plate 8.3., p.33).

> De nombreux cailloux du remplissage würmien de l'Hortus sont craquelés en surface. Leur face a l'aspect d'une mosaïque de fragments anguleux séparés par des fissures de 1 à 3 mm. .... (Miskovsky 1972:112)
>
> [Trans.: Numerous stones in the würmian fill at l'Hortus [Hérault] have surfaces covered in cracks. Their surface appears as a mosaic of angular fragments, separated by fissures of 1 to 3mm [depth or width, both being equally compatible with the photographs] ....]

This description is reminiscent of that given by Laville.

Miskovsky continues:

> L'origine de ces craquelures n'est pas facile à expliquer. Nous avons imaginé deux hypothèses:
> - craquelures par dessiccation,
> - craquelures par le gel.
> Un caillou corrodé en surface, et placé dans un milieu très sec, peut se craqueler et les bords des différents fragments se recourbent alors vers l'extérieur. Il est difficile d'expliquer par un tel phénomène le craquellement des cailloux de l'Hortus. Ceux-ci en effet, ont été découverts dans un fossé situé à l'entrée d'une grotte, donc dans un milieu qui a dû toujours être plus ou moins humide.
> Les craquelures de ces cailloux paraissent avoir été provoquées plus vraisemblablement par l'action du gel. Un tel phénomène a pu être réalisé expérimentalement en laboratoire sur des cailloux en calcaire marneux et nécessite des alternance de gel et de dégel rapprochés, et la présence d'eau. Il faut cependant signaler que seule, la zone superficielle des cailloux de l'Hortus est craquelée et que dans les cailloux expérimentalement craquelés par le gel, ceux-ci sont brisés dans toute leur masse. Il est alors possible de penser qu'à l'Hortus, l'action du gel n'a eu d'influence que sur la zone périphérique et poreuse des cailloux qui s'imbibe plus facilement en eau. Ceci est très vraisemblable car nous avons constaté à l'Hortus que seuls, les cailloux corrodés étaient craquelés par le gel, tandis que les cailloux non usés ou peu usés l'étaient faiblement. (ibid., p.112)
>
> [Trans.: The origin of these cracks is not easy to explain. We have thought of two hypotheses:
> - cracking by desiccation,
> - cracking by frost.
> A stone with a corroded surface, and placed in a very dry context, may crack and the edges of the different fragments then curl outwards. It is difficult to explain the cracking of stones at l'Hortus by such a phenomenon. In fact, they were discovered in a rift sited at the entrance to a cave, thus, in a context which must always have been more or less damp.
> The cracking of these stones appears more probably to have been caused by frost action. It has been possible to produce such a phenomenon experimentally in the laboratory, using stones of marly limestone; rapid freeze-thaw cycles and the presence of water are necessary. However, we must point out that only the superficial zone of the stones from l'Hortus is cracked, whilst the stones cracked experimentally by frost are shattered right through. It is therefore possible to believe that, at l'Hortus, frost action has affected only the peripheral, porous zone of the stones, which takes up water more easily. This is very likely because we have noticed that, at l'Hortus, only corroded stones were cracked by frost, whilst unworn or slightly worn stones were [only] slightly cracked.]

Miskovsky's discussion is most interesting, especially the reference to the attempt at replication of the cracking. It is a great pity that so few experiments of this sort, with a specific problem in mind, are undertaken in cave sedimentology.

However, the present author attaches great importance to the fact that cryoclastis tends to shatter rock right through or, if not, to produce nothing more than a few hair-line cracks. The author has carried out his own freeze-thaw experiments with various types of limestone, of varying degrees of alteration; gaping cracks of the type recognised by Miskovsky and Laville have never occurred. Nor has such a phenomenon ever been reported in the considerable literature on experimental cryoclastis, to the author's knowledge, although, as intimated by Miskovsky, such forms are known from hot desert environments (cf. Ollier 1969). Only one of the author's sites (Sun Hole, Mendip; cf. p.918) has produced such cracked or fissured, even 'crazed', limestone clasts (as well as similarly cracked crystalline calcite concretions), and this in a context with a clayey and moderately organic-rich matrix and various faunal elements suggesting an interglacial origin. The most affected clasts are visibly altered and the sides of the cracks, as seen in cross-sections of stones, diverge as marked curves: the limestone is not only cracked but actually deformed. Similar cracking is quite common in weathered bone and coprolites in many sites. Finally, cracked limestone from Sun Hole, when subjected to only ten freezing cycles (14h at -10°C, 10h at 20°C) was reduced almost completely to its component fragments. All circumstantial evidence points to a combination of dominantly chemical processes (slaking, hydration, solution, etc.), rather than frost, as the primary cause of this phenomenon. This would explain both Miskovsky's observations and the fact that Laville never reports 'fissured' and 'frost-shattered' stones in the same layer.

Laville is rather mysterious with respect to the conditions which would lead to the formation of 'plaquettes'. Under certain

conditions of temperature, humidity, aspect and lithology, cryoclastis of *in situ* rock leads to the formation of 'plaquettes' - that is all we are told. However, in his Plate 3 (1975:17), Laville provides two admirable photographs of impressive quantities of 'plaquettes' demonstrably forming under present day conditions in S.W. France. He also references two other Périgord sites where 'plaquettes' are now being detached (ibid., p.13). All these sites are damp and shady; two face north (Le Ruth and Sourzac), one east (Laugerie) and one south-west (La Faurélie). The rocks involved would appear to be weakly bedded calcarenites. The 'plaquettes' tend to fall in the winter half of the year.

Such 'plaquettes' are easy to find all over the lowland limestone areas of Britain, especially in damp situations, maintained either by abutting sediment or by algal and moss growths. Upon visiting the site, the author was not surprised to find that the 'engraving of a cervid head', recently claimed from the Wye Valley (Rogers et al. 1981), is entirely composed of partially detached 'plaquettes' in the damp basal notch between rockface and sediment. Frost action may well play some part in the final detachment of 'plaquettes', but whether cryoclastis is always the main cause of the initial fracturing, as suggested by Laville, seems open to debate. The very fact that one surface of the 'typical plaquette' is said to be more altered than the other means that a degree of alteration must intervene between the detachment of elements from the same position on the rockface. In any case, Laville's examples of contemporary 'plaquettes' imply that it is not automatically necessary to invoke more severe frost action than is now occurring.

Having described 'typical plaquettes', Laville then

changes the practical definition to a mere ratio between thickness and width. The reason for this is given as the alteration of the originally fresher surface, although why 'plaquettes' should be vulnerable to alteration and not 'frost-shattered' stones (supra) is unclear. The present author is not willing to accept the automatic equivalence of any platey limestone clast and those clasts discussed above, nor the even more extreme inference that platey limestone clasts are necessarily a direct result of frost action. At Pontnewydd (Clwyd; partially recrystallised calcite mudstone, Carboniferous), Pin Hole (Derbyshire; granular magnesian limestone, Permian) and Pixie's Hole (Devon; massive limestone, Devonian), large flat clasts may be observed that have demonstrably spalled off cave walls already buried by sediment. There is absolutely no sign of cryoclastis in the relevant sediments and limestone particles larger than sand are either extremely rare or demonstrably imported by mass movement. The spalls from the cave walls have migrated out into the surrounding sediment where many of them have then split into even thinner plates. Most of these clasts are still vertical, but some have rotated and then split; this could only have occurred after migration away from the wall since some at least of the component plates would have become dissociated otherwise. If these clasts can split, so can any clast, whether or not it was originally derived from buried bedrock. Some limestones (and minor, but common, stratigraphic units within limestone sequences) are thinly bedded and will automatically form platey clasts, no matter what the shattering process.

Miskovsky (1972:107) also recognises 'plaquettes' at l'Hortus. Two types are of interest here: 'thin plaquettes',

with two main surfaces parallel and the thickness less than a quarter of the maximum width; and "plaquettes gélives" (frost-riven), which are 'very thin' (not quantified) and have one sharp edge at the convergence of the two main surfaces. No reason for suggesting frost shattering in connection with the latter type is given. However, the requirement of one sharp edge, and thus a thin wedge shape, suggests that Miskovsky sees these clasts as the result of the removal of spalls from the rockface, one or two at a time. Miskovsky discusses the results of counts of these types from l'Hortus only briefly, but the point which the present author finds of interest is that the percentages of both forms of 'plaquettes' in the holocene levels are slightly higher than the percentages in Layers 26-31, considered to be a cold and slightly humid phase during the Wurm II, and also corresponding to the maximum of cool biotope birds and one of only two brief appearances of reindeer at this site.

In a later publication, Miskovsky concludes:

> Le pourcentage de plaquettes, prismes et polyèdres ne paraît pas être toujours lié aux influences climatiques. Les résultats obtenus ne donnent en général aucun renseignement pour la Paléoclimatologie. Cependant le délitage par le gel de certaines roches gréseuses ou molassiques entraîne la présence de nombreuses plaquettes. (1974:39)
>
> [Trans.: The percentage of 'plaquettes', prisms and polyhedrals does not always appear to be linked with climatic influences. In general, the result obtained give no palaeoclimatic information. However, the desquamation by frost of certain sandstones [textural not compositional] or molassic rocks [tectonic not textural or compositional] produces numerous 'plaquettes'.]

Miskovsky had earlier (1972) differentiated 'thick plaquettes' (two main surfaces parallel, but thicker than a quarter of the maximum width) from the two types already discussed above. The lack of differentiation in the statement made in 1974 therefore renders the exact meaning unclear.

Nevertheless, it would appear that Miskovsky is suggesting that lithology plays a particularly important role, and the word "délitage" may well have the even more restricted sense of 'splitting along bedding planes'. Three sites are singled out as having particularly important accumulations of 'plaquettes', assumed to be the result of cryoclastis. The geology of the bedrock of these sites is always described mainly in terms of chronostratigraphic units, but a few details of lithology may be extracted: Baume des Peyrards (Vaucluse), very thin bedded and fissile calcareous molasse; l'Hortus (supra), relatively thin bedded and well jointed zoogenic, marly or gravelly limestone; Le Lazaret (Alpes-Maritimes), marine conglomeratic fill of a fissure in dolomitic limestone.

Miskovsky's opinion that prisms, polyhedrals and (at least) 'thick plaquettes' do not give climatic information contrasts with suggestions by Brunnacker for the site of Crvena Stijena (Montenegro, Yugoslavia):

> Die Frostverwitterung.
> Die Theorie der Untersuchung geht in sehr vereinfachter Weise davon aus, dass vom Höhlendach durch Frost gelockerte Gesteinsstücke mit verhältnismässig einfachen Formen abbrechen, z.B. dem 6-, 5-, 4-Flächer angenäherte Körper. Beim Aufprall am Höhlenboden vermögen die Steine noch weiter zu zerbrechen. Je länger die Stücke am Höhlenboden liegen bleiben, desto mehr steigt die Chance für eine weitere Zerlegung in kleinere Teile nicht nur mit etwa gleicher Flächenzahl sondern auch dafür, dass durch zusätzliche Abwitterung an Ecken und Kanten weitere Flächen an den einzelnen Körpern entstehen (vgl. K. Brunnacker & R. Streit 1965). Geringe Werte möglicher Bruchflächen zeigen also auf geringe Frostverwitterung am Höhlenboden, hohe Werte dagegen auf eine stärkere derartige Aufbereitung. Voraussetzung ist allerdings, dass dafür Frostwirkung und Feuchtigkeit am Höhlenboden wirksam werden. Das Ergebnis ist für die Fraktion 6-10 mm.Ø ... zusammengefasst. (1966:47)
>
> [Trans.: Frost Weathering.
> The theory of the analysis is based on the simple premise that stone fragments, detached from the cave roof by frost, will break away with relatively simple forms, that is, they will approximate to bodies with 6, 5 or 4 surfaces. The stones may also break further upon striking the cave floor [or sediments]. The longer the fragments remain lying on the cave floor, the greater is the chance of a further breakdown into smaller pieces, not only perhaps with the same number of surfaces but also with additional surfaces on individual elements, formed by

> further [mechanical] weathering near edges and corners (cf. Brunnacker & Streit 1965). Therefore, low values of possible break surfaces indicate little frost weathering of the cave sediments, whilst comparatively high values show that this process was stronger. It is assumed, of course, that frost action and moisture were present in the cave sediments in order that the process might occur. The results are summarised ... for the 6-10mm diameter fraction.]

The cave is developed in a thick lens of dolomite and dolomitic limestone, with much of the scree derived from the reef limestone which forms the roof and cliff-face above. This approach to clasts shape is simple and therefore attractive. However, lithology will still play a dominant role, since the logical prediction that the number of surfaces on some clasts will increase significantly with increased shattering will translate into different absolute figures, depending upon whether the limestone is fissile or isotropic (just as easily broken in any direction). On the other hand, it ought to be relatively easy to discover how a given limestone sequence will behave, either by freeze-thaw experiments or simply by crushing samples. One practical problem is where to place the dividing line between minor edge damage and small facets worthy of counting. Perhaps as a constraint of sample size, Brunnacker uses very small clasts for this analysis (a typographical error seems unlikely, since the class limits are mentioned several times elsewhere in the text), clasts with surface morphology that will be very quickly obscured by chemical alteration or physical edge-rounding. It is therefore surprising that he apparently had no problem in counting surfaces, even on the most heavily corroded and rounded clasts from Crvena Stijena. When all is said and done, given a 'sympathetic' lithology and negligible surface degradation, counts of the number of surfaces, or of angles (cf. Riek 1957), on individual clasts may furnish some idea of the intensity of shattering. Nevertheless, we are no nearer solving the problem

of how to distinguish frost shattering from any other type.

The conclusion that the present author draws from the above discussion of particle size and shape is that these properties give only very rough estimates of the intensity and nature of cryoclastis in limestone. There are simply too many factors which might influence scree production and modification for the 'saw-tooth' patterns, which usually result when the values of these textural parameters are graphed for a given sequence, to be considered as detailed indicators of frost activity. A statement by Lautridou, which should be valid at least for secondary cryoclastis, sums up the position:

> L'interprétation paléoclimatique des formations périglaciaires doit être faite avec prudence: l'intensité du gel ne semble pas déterminante (*); le degré d'humidification et la répétition des cycles gel-dégel constituent les deux facteurs essentiels. Il importe avant tout d'étudier le comportement de chaque faciès en le replaçant dans un des cinq types définis (craie - calcaire crayeux - calcaire lithographique - calcaire crystallin - calcaire perméable). En dépit des nombreuses recherches des physiciens, les mécanismes très complexes du gel dans les roches ne sont pas encore parfaitement connus.
> (* Cette loi est valable lorsque le temps de gel est suffisamment long pour geler complètement les blocs; un gel court et peu intense, n'atteignant pas le coeur des cailloux, s'avère peu efficace ("essai islandais" de TRICART).)
> (1971:79)
>
> [Trans.: Palaeoclimatic interpretation of periglacial deposits must be carried out with caution: the intensity of freezing does not appear to be determinant (*); the degree of humidification and the repetition of freeze-thaw cycles constitute the two essential factors. Above all, it is necessary to study the behaviour of each [rock] facies, placing it in one of the five defined categories (chalk - chalky limestone - lithographic limestone [calcite mudstone with slight fissility] - crystalline [massive] limestone - permeable limestone [calcarentites, etc.]). Despite the numerous studies by physicists, the extremely complex mechanisms of freezing within rock are not yet perfectly understood.
> (* This rule is valid when the freezing time is sufficiently long for the blocks to be completely frozen; a short frost of low intensity, which does not reach the core of the stones, is not very effective (the "icelandic trial" of TRICART).)]

It is often suggested that textural parameters should provide relative values, which will be reliable for comparison within a given sequence. The author is convinced that scree production is one of the processes which is most strongly

affected by the past history of a cave site. This 'memory' effect, which may often involve 'events' that took place hundreds or thousands of years before the formation of a given scree body, makes a nonsense of any assumption that the exact nature of the scree is an instantaneous response to the prevailing climate. In addition, there are processes that continue over hundreds or thousands of years which progressively weaken the bedrock. These 'events' do not include major geomorphological phenomena (such as the general processes of denudation), that take place on a scale of tens or hundreds of thousands of years, nor geologically ancient phenomena that have produced the observed lithology and structure of the limestone bedrock. The 'events' of interest here comprise the much less obvious effects of the complex interaction of many physical and chemical processes which prepare the limestone for eventual shattering. Lautridou (1971) has suggested fatigue, caused merely by temperature oscillations, as a factor in the shattering of massive limestones; microfissuring eventually reaches a density at which sufficient water can penetrate the rock for true cryoclastis to begin to act. Bertouille (1975) was unable to shatter several rock types by experimental cryoclastis unless unnaturally rapid and extreme freeze-thaw was applied. Since these rocks are known to have produced impressive pleistocene screes, Bertouille also suggests that fatigue must be an important element. Similarly, the present author was able to produce only a few sand and silt grains by subjecting c.8cm unweathered clasts of calcite mudstone from the bedrock at Sun Hole (cf. p.918) to 200 frost cycles (14^{h} at -10°C, 10h at 20°C, in a tray with 2cm of water). Yet clasts of this rock, of various shapes and sizes, are extremely common in sediments, of various ages, in Cheddar Gorge. Any rock

that is not frost-susceptible in the fresh state must surely become increasingly vulnerable with the cumulative effects, not only of 'thermal' and other types of physical fatigue, but also of various sorts of chemical alteration. It is difficult to see how such effects could be accurately duplicated in the laboratory, especially on the scale of the outcrop, if considerable periods are involved in nature. It is easy to induce fatigue experimentally, but the degree and nature of fatigue are unalterably a function of time and there is no satisfactory method of producing an accurate 'scalar model'. One of the problems with many British fine grained limestones is that they often include extremely thin veins of calcite. Apart from introducing a structural flaw, this calcite often degrades much faster than the limestone itself, causing the rock to split up into angular fragments that do not necessarily give any obvious clues as to whether physical or chemical processes were involved. It should not be assumed that all significant chemical processes are inactive during 'glacial' periods.

A statement by Bowen, concerning open-air sequences along the Gower coast, is of interest here:

> Basal colluvial silts indicate the onset of cold conditions when sheet washing and soil erosion occurred on a more or less treeless land surface. Following the removal of fine grained regolith in this way, subsequent frost action detached larger particles which accumulated at the cliff base as a 'blocky' litho-facies. This material had been subject to preparatory chemical weathering during the previous temperate episode. Exposure of fresh unweathered bedrock to frost action produced the finer head litho-facies. [...]
> [...] It has been shown, however, that head sequences fining upwards are a consequence of a cyclical denudation of adjacent rock slopes .... (1977b:233-4)

Laville (1975:22) also presents a similar sort of argument, though with different details, when discussing the deepening of rockshelters immediately after interglacials and interstadials. For the Gower sequences, Bowen is also proposing

a long time scale, with the fining upward sequences representing a whole glacial period (the Devensian). Nevertheless, it is clear that Bowen is suggesting a most significant 'memory effect'. Is it possible that shorter cycles might occur and, if so, what would they look like?

The entire devensian sequence at Bacon Hole (p.1054) is composed of fining upward cyclothems, starting with gravel of various calibres and sometimes ending in a phase of carbonate cementation; each unit is relatively thick and probably represents several centuries (cf. similar examples given by Miskovsky 1974:39). At Sun Hole (p.909), the deposits of the Late Devensian comprise discontinuous fine gravel lenses floating in a finer, but badly sorted, sediment; the regular alternation of finer and coarser material seen along any vertical line probably represents a rhythm of a few decades. The logical end member of this series is the occurrence of grèzes litées (sensu stricto; cf. Guillien 1951; Lautridou 1971). These are comparatively thick beds of very fine gravel, or even coarse sand, alternating with thin beds of finer material, a pattern considered (by most authors except Guillien) to be the result of seasonal, or even diurnal, temperature oscillations (cryoclastis) in a snow-rich environment (cf. section 10.1.). Note that grèzes litées are not necessarily equivalent to éboulis ordonnés, 'stratified scree' or 'sorted talus' (contra Butzer 1971; French 1976), all terms that have been used to cover a wide range of structured slope phenomena, from true grèzes litées to cyclothems of the Bacon Hole type and also sometimes including rhythmic mass movement deposits. Grèzes litées have apparently been recognised at the shelter of La Ferrassie (Dordogne), presumably in Layers B to D (pers.comm. from H. Laville, reported in Texier

& Thibault 1976, but not confirmed in Laville et al. 1980).

It is clear that there can be little or no 'memory effect' involved in the formation of grèzes litées. However, moving back up the scale, the Sun Hole rhythms may be due to the need for a particular patch of bedrock to receive a threshold level of physical and chemical preparation before the next gravel lens is formed immediately below. Breakdown might be more or less continuous in the site as a whole, but the active locations would 'wander' about the walls and roof. At Bacon Hole, the level of preparation is perhaps, at least in part, climatically controlled, with phases of preparation and freeze-thaw becoming more temporally distinct. However, a very careful examination would be necessary to identify any possible inclusion of Sun Hole rhythms (or, in more open situations than Bacon Hole, of grèzes litées), together with any possibly significant lithologic changes. Note that, even after the most careful study, we are more likely to be able to come to conclusions about the phases of alteration than about the phases of cryoclastis. Nevertheless, the study of sequences, especially those showing cyclicity of some sort, would seem to be the best approach for the future, although the present author remains pessimistic about the chances of recovering anything more than very general indications concerning cold climates in the absence of the types of structures discussed in section 8.2. It should not be forgotten that 'sequence' has both a vertical and lateral aspect; Guillien (1970) has shown how important the study of lateral variations can be in the understanding of the origin of scree in caves.

9. MASS MOVEMENT

9.1. Generalities

9.1.1. It is the author's belief that, in the context of the European Quaternary, mass movement of one sort or another may account for a highly significant proportion of cave deposits. Within particular sequences, mass movement may even be the dominant process.

This belief, which will be justified by many examples throughout the text, calls for a detailed examination of what is an extremely complex subject. The deposits resulting from mass movement have rarely been examined in detail in the cave situation. Such material, when recognised, has usually been called either 'collapse and rockfall deposits' or 'solifluction deposits'. For obvious reasons, exact observations of the processes involved in the faster types of mass movement are almost totally lacking. Consequently, it will be necessary to draw heavily upon research carried out either in the laboratory or in open-air sites.

In the following, the distinction between process and the resulting deposits is of the utmost importance. There is by no means a simple one-to-one relationship between the two.

9.1.2. Definition

A relatively consistent distinction exists in sediment transport

> processes between those in which large quantities of sediment move together in close grain-to-grain contact and those where transport is mainly in the form of single grains dispersed through a fluid transporting medium. The former are often called mass movement processes . . . . The distinction between mass movement and flow processes is to some extent genetic, and is one which is extremely convenient for discussion purposes. But it is important to remember that, on close examination, sharp distinctions between the two broad process-groups merge into a gradual transition. The continuum between flow and mass movement processes makes a purely rational treatment of transfer processes rather difficult and some genetic relationships have been sacrificed to expedience [here]. In particular, flowslides [cf. debris flows, infra], some of which can justifiably be argued to be dispersed flow processes, are dealt with [here]. Also the process of rockfall is treated as a mass movement simply because it is difficult to classify in any other way. (Statham 1977:70)

> The general term mass movement is applied to those processes that involve a transfer of slope-forming materials from higher to lower ground, under the influence of gravity, without the primary assistance of a fluid transporting agent. They merge imperceptibly with processes in which transporting media, such as air, water or ice are involved and which are generally called mass-transport processes. The movements may be slow or rapid, shallow or deep and include one or more of the mechanisms of creep, flow, slide or fall. (Brunsden 1979:130)

The term mass movement will be used in the following as defined above, but will also include Brunsden's "mass-transport" processes.

9.1.3. Classification

The factors of interest to us may be grouped under four main headings:

Time scale. Is a given deposit due to a more or less instantaneous 'event', a series of disjointed 'events', or a more normal pattern of gradual sedimentation? This question involves the actual recognition of mass movement deposits.

Causes. What factors are responsible for the initiation, maintenance and cessation of movement? Are climatic factors important or are they only involved as part of the general environmental system?

Effect on older material. Are the movements erosive,

destructive or deformational?

Internal characteristics. What are the carrying and sorting capacities, the potential for damaging or mixing various components, the distance from the main source, the direction of movement, etc.?

As a general proposition, we may say that, from our point of view, mass movement is an unwelcome intrusion. The normal suppositions concerning stratigraphic superposition and association of objects in a single deposit must be suspended. Originally well ordered temporal and spatial information may now be distorted, fragmentary of irretrievably lost. We therefore require a classification system for the various types of mass movement that reflects our concern.

Many classification systems have been proposed in the literature. Naturally, the details depend upon each author's own interests and are variously based upon considerations of the physics involved, speed of movement, lithology, geography, geomorphology, climate, etc. A disturbing factor is that different authors may use the same term to describe very different phenomena or, conversely, one author may use the same term to cover a wide variety of phenomena. For instance, what is one author's 'mudflow' may be another's 'earthflow', 'mudslide', 'debris flow', 'flowslide', 'lahar', 'debris avalanche', etc. Although every author may define, more or less precisely, what he means by 'mudflow', if one takes the lowest common denominator over a number of publications, one finds that 'mudflow' is just about any kind of mass movement involving wet material.

Consequently, although a considerable effort has been made here to use terms only with their most conventional meanings, it will only be through precise, and sometimes rather arbitrary,

definitions that accurate communication can be achieved. A complex formal classification is of no use to us. Rather than finding labels for each possible variant, we will concentrate upon how different processes produce different, or similar, combinations of sediment parameters. It is these parameters, especially those defining structural and geometric features, that we must observe in fossil deposits.

For our purposes, we may envisage three main groups of mass movement phenomena:

i. Material falls, slides or is otherwise disrupted, without significant lateral movement away from the source area. We may often hope to be able to examine the undisturbed source.

ii. Material moves away from the source area at speeds which would require long-term measurement in order to demonstrate movement. We may often hope for only moderate information loss.

iii. Material moves away from the source area at speeds which would be noticeable to an observer over quite a short period. We must expect maximum information loss.

These groups are of course only qualitative but most observed phenomena fit quite easily into the framework.

9.2. Mass Movement With Little Lateral Displacement

9.2.1. It will be assumed in this section that the undisturbed source is available for study. We need therefore only be concerned with information regarding the causes and effects of movement. Note that, in company with most reference texts (supra), some phenomena involving movement of individual particles rather than mass movement (eg. rockfall) are covered here.

9.2.2. The Collapse of Hard Rock

We are concerned with the collapse of limestone, often a well bedded, jointed and faulted rock type. The various causes of limestone collapse, a subject of great importance to cave sedimentology, are, for that very reason, discussed elsewhere (see p.284). The process is largely spasmodic, with individual particles acting separately. Classification is usually based on scale, which, leaving environmental/climatic factors aside, is a function of topography and lithology.

The smallest scale of rock collapse is granular disintegration, in which individual mineral grains or other small structural elements (eg. fossils) fall independently. Although this must happen to some extent with all limestones, it is only with particular lithologies that the effect is really noticeable. Farrand (1975) reports the use of a simple sediment trap installed to catch sand at the Abri Pataud (Dordogne) and monitored over 3316 days (just over nine years). The results averaged 182.5 g/yr over an area of 1485 cm^2. Farrand calculates a sedimentation rate of 90 cm/ky, not allowing for compaction. Not surprisingly, the rock is a calcarenite (Coniacian, Upper Cretaceous). At the entrance to the Grotte de Verlaine (Prov. de Luxembourg, Belgium), the present author has observed over three metres of holocene deposits, the matrix of which consists almost entirely of carbonate sand. This is a dolomitic limestone (Devonian) with a truly saccharoidal texture. Some care must be taken in the interpretation of carbonate sand, as, in addition to free-fall from exposed rockfaces, such sand may also be derived from post-depositional decomposition of larger clasts (cf. Bacon Hole, p.1041; Brunnacker & Streit 1965).

The recognition of granular disintegration not merely as a possible source but as an actual primary source for a given deposit is far from easy. Due to the scale, such fine material will almost certainly be reworked. Consequently, diagnostic sediment bodies have little or no chance to form. Even at the Grotte de Verlaine, where granular disintegration is extremely active today, the deposits are bedded at angles well below the theoretical angle of repose for such materials, indicating that other processes (eg. creep, wash or even wind) constantly redistribute the sediment. If this sediment were to be examined at some time far in the future, it might be difficult to tell whether it represented contemporary granular disintegration or, perhaps, much older material which had later been washed off the cliff top (or out of the cave itself which contains substantial deposits of this type) during a period when granular disintegration was not in fact operative.

When fragments appreciably larger than individual grains collapse, the process is rockfall. This term will be used here in the sense recommended by Statham:

> Rockfall is a small-scale process operating on almost all rock walls in some degree, regardless of the over-all stability of the slope. It is a process by which relatively small blocks of rock are detached from the cliff face, falling freely through space to the foot of the cliff. Rockfall can take place only under certain conditions: small joints and fissures are essential and the rate of joint-opening must be substantially in excess of the rate of grain detachment. (1977:87)

There is no exact upper size limit for this material, but a general distinction will be made between these rocks and large units (boulders, roof-slabs, etc.). It would be superfluous to cite examples of this type of material. Rockfall may be considered to be at least the first process in the chain which provides massive quantities, and often the dominant component, of cave sediment. However, we are here concerned with distinct

deposits which result more or less exclusively from this process. Such deposits are generally termed scree. The word talus is also legitimate but it should be noted that many authors, especially those involved in cave research, may use this term to indicate any broadly coarse, sloping deposit, no matter what its origin. The French word éboulis (plural invar.) has sometimes been used in English and is an exact synonym of the word 'scree' as used here.

If scree production is slow compared with other sediment input processes no distinctive rockfall deposits will result. Nevertheless, useful information can be retrieved concerning the likelihood of lateral movement or the origin of concentrations of stonier material. Fedele (1976) has noted the links between cave morphology and scree production. When rockfall becomes the dominant process, recognisable deposits will be formed. A characteristic of such deposits will be that, whether or not a fine matrix is present, the coarse fraction will be more or less autonomous; it is the limestone elements themselves, in frequent contact one with another, that support the deposit, not the matrix. Care should be taken to differentiate, if possible, between such deposits and residual coarse deposits from which the original matrix has been removed (see p.442). Most primary screes will have a rather loose and unstable fabric, with packing far from the theoretical maximum.

There are various styles of scree bodies depending upon the morphology of the source and of the receiving area, and also upon the time scale involved.

True collapse deposits originate either from pre-existing scree or from heavily fractured bedrock. Their emplacement is 'instantaneous' and their fabric will be more or less chaotic.

The geometry of the deposits will depend upon spatial relationships with the source but they will generally occur as rather irregular cones or fans. If material collapses directly onto a significant slope, some fabric features may result that will resemble those developed by more continuous rockfall (infra). If continuous rockfall occurs in a restricted space (eg. a vertical shaft) the resulting deposit may be rather chaotic and difficult to distinguish from a collapse, although seasonal variation in parameters such as clast size may create banding that would be most unlikely to have resulted from a single collapse event (see p.310).

Scree bodies that have formed over longer periods have certain diagnostic features. Since, by our definition, scree production must outweigh other sediment input processes, it follows that rock particles will constantly be falling upon existing scree. That impact is an important factor is supported by Gardner (1969), who observed that movement of marked stones increased towards the top of scree slopes, directly under the source rockfaces. This will result in comminution by percussive fracture. Also, once a particle is incorporated into loose scree, it may become subject to very high point loading from overlying material; the static 'pressure' fracture thus developed is mechanically identical to dynamic percussive fracture. Some limestones, especially the finer grained types referred to as calcite mudstones or 'chinastones', acquire conchoidal fracture surfaces that are quite different from surfaces due to flexure or shear. These surfaces may be recognised in ancient deposits (cf. Sun Hole, p.918). It is surprising that so few limestone flake 'artefacts' have been claimed, given that such natural particles often carry all the stigmata of struck flint.

In large cave chambers, scree may be produced more or less uniformly from wide areas of the roof. If, in addition, the cave floor is approximately horizontal, the scree will build up over the whole floor with little or no organised fabric (cf. Bacon Hole, p.1072). The tabular geometry of such a deposit would suffice to distinguish it from a simple collapse. Such a situation, producing a widespread featureless deposit, is probably rather rare. Usually rockfall occurs more intensely at preferred sites, such as cliff or wall faces (linear source), cave entrances, in chambers under feeder shafts (point source), or beneath structurally unstable zones in the rock. At first scree builds up directly under the source, but this cannot continue for very long and a slope, away from the source, is initiated. As rockfall continues, a slope approximating to the angle of repose for that material is achieved. This angle varies with clast morphology and packing but it is normally very high, in the range 30-40° (King 1966; Carson 1971; Embleton & King 1975; Statham 1976; Brunsden 1979). King (1966) suggests that "angular" (blocky?) fragments may achieve higher stability, and thus higher slope angles, than platy fragments. Similarly, Statham (1976b) suggests that flatter fragments are more stable than spherical fragments. This characteristic slope is of the utmost importance. A study of the general archaeological and palaeontological literature concerning cave sites will reveal countless examples of thick deposits identified as primary screes that are developed at much lower angles. In the author's opinion, many of these cannot be primary screes; the lower angles must reflect additional movement (probably slow mass movement, <u>infra</u>) which may have had important effects on included artefacts or fauna.

Most observers report quite efficient size sorting,

coarsening downslope, in screes (cf. King 1966; Bones 1973; Embleton & King 1975; Statham 1976b; Loughran & Loughran 1979). The fact that impact comminution is concentrated at the top of the slope must contribute to this pattern, but it is also clear that true sorting occurs. Statham (1976b) attributes sorting to the fact that larger particles can overcome the roughness of the scree surface better than can smaller particles. This is also, of course, a kinetic factor (Washburn 1979), larger particles having greater momentum (infra); large blocks may even 'bounce' well beyond the scree body itself. Size sorting may therefore depend upon the height of the feeder cliff and 'reverse sorting', with larger material at the top, may be expected with very low cliffs involving practically no fall (cf. Rapp 1960).

If a scree slope, or indeed any steep slope where rockfall is contributing material, is covered by snow, coarse particles may slide down the slope and accumulate as a ridge at the base. Large scale features of this kind are called protalus ramparts (cf. Embleton & King 1975:140), but analogous small scale features should be recognisable. Care should be taken to differentiate such primary features from others produced by slow mass movement. Protalus ramparts are coarse, usually lacking a fine matrix, and they do not have an oriented fabric (Washburn 1979). They are often arcuate in plan.

There is a tendency for the longitudinal profiles of scree slopes to be rectilinear over most of their length, although there may be a greater or lesser concavity at the base, often due to other transport processes (wash, slow or fast mass movement, etc.)(French 1976; Brunsden 1979; Washburn 1979). However, a concave profile may develop, with even better sorting, if rocks are falling from all levels of a high cliff face (cf. Rapp 1960).

Statham (1976b) interprets this feature as being due to the exponential growth in momentum (i.e. acceleration due to gravity) as greater and greater fall heights are involved. This theory would appear to be sound as individual cases can be modelled quite accurately. However, it is difficult to give a general definition of 'high' since many factors are involved. Certainly, the present author has observed approximately straight profiles under limestone cliffs in the Pennines of up to c.30m in height (cf. Sparks 1972, plate 4, p.71). King (1966) notes that scree in gullies often has a convex profile, presumably due to added support by friction against the gully walls. It is possible that this effect could also occur in narrow cave passages. Note, however, that scree directly below a gully will usually have a concave profile, due to concentration of drainage through the gully and consequently increased wash and mass movement.

Preferential fabric would appear to be rather variable in scree deposits (Washburn 1979). Many authors make no mention of such features as particle orientation, which suggests that fabric may sometimes be rather irregular. Statham (1976b) notes predominant orientation of long axes with the slope, due to sliding, but also some orientation perpendicular to the slope, due to rolling. He also suggests a slight shape sorting effect, produced by more spherical particles rolling further downslope (supra), but Loughran and Loughran (1979) observed no significant change in sphericity in their sample. Statham (1976b) reports weak imbrication. King (1966) notes the higher stability of screes showing reverse imbrication; it is not at all clear how such a fabric might develop. Other features, such as 'heavy end' orientation, have not, to the author's knowledge, been reported. From general theory, it should follow that preferential fabrics

should be highly dependent upon the shape and particle size distribution of the source rockfall material. It has been suggested (McSaveney 1972) that the degree of fabric development is related to surface roughness. Thus coarser screes, or areas of a scree slope, should show less well developed fabric.

Mechanical rounding of scree, apart from the removal of fragile protuberances and edge crushing, is at a minimum, since sliding appears to be the dominant movement (Washburn 1979). Larger intrusive elements that are brittle (eg. flints) or relatively soft (eg. bone) will suffer extensive damage; smaller elements may penetrate through the openwork fabric to levels far below their own chronohorizon.

Large boulders or roof slabs constitute a logical continuum with finer rockfall deposits but, for many reasons, they are better treated separately. Such elements will be called 'slabs' in the following discussion, merely for ease of reference.

The size and shape of slabs are primarily a function of major planes of weakness (joints, faults, bedding planes, etc.) in the rock. This is demonstrable, for instance, when slabs show a surface composed of an intrusive lithology, such as calcite or other vein minerals (cf. Joint Mitnor, p.806). Large scale failure in hard rocks takes several forms (eg. plane, wedge or toppling failures, rockslides, etc.; cf. Statham 1977; Brunsden 1979) but the details need not concern us here. Since the intervals between major planes of weakness are usually markedly greater than the intervals between minor discontinuities, one might expect clear separation in the size ranges of slabs and scree. However, it would be almost impossible to prove this as the required sample volume would be enormous. Certainly, in any given deposit, there is often an apparent difference of some

magnitude. Note that even when sound slabs fall, the impact may often set up a network of fissures. Although these may not be obvious immediately after the fall, weathering processes and various stresses acting after burial and over long periods may reduce the slab to material very like scree. Careful observation will, however, reveal how this material 'refits' (cf. Bacon Hole, p.1054). If the fissured slab is attacked before burial, spreads of true scree may result.

The main importance of slabs to cave research lies in the fact that the fall of a slab will change the sedimentation pattern to a greater or lesser degree. The presence of slabs may indicate the opening (cf. Bacon Hole, p.1063) or closing (cf. Joint Mitnor, p.815) of a major input route for sediments. Slabs may change the micro-environment of caves. Slabs often delimit new basins of sedimentation. Indeed, very large slabs may even form a new 'cave' or shelter (cf. the sites of Laugerie, Dordogne). Consequently, it is extremely important that the exact stratigraphic position of slabs be recognised, information which is not always easy to recover with large slabs only exposed in a single section. It should not be assumed that even the largest slabs have not moved since their fall. Slabs may sink into older, incompetent strata under their own weight, a process which may be accelerated if the older deposits become subject to solution or removal of fine material. Renewed mass movement may transport or disturb slabs of impressive dimensions.

9.2.3. The Collapse of Indurated Sediment

Cave sediments may become indurated due to the input or mobilisation of a cement, usually a combination of carbonates

but sometimes including phosphates and complex (hydr)oxides of iron, aluminium and manganese (see p.491). In this category, we may include various breccias, conglomerates and stalagmitic floors. If these deposits are later eroded, hanging remnants may be left adhering to the cave walls; they may then act as a source for new sediments (cf. Three Holes Cave, p.776; Sun Hole, p.947). Input may vary from the fall of individual particles to the collapse of large aggregates. Since these hanging deposits may contain artefacts, fauna, environmental or chronological indicators, etc., it is of the utmost importance that they be examined (no matter how difficult they may be to reach!), both in their own right and also as potential contaminants for younger deposits.

9.2.4. The Collapse of Unconsolidated Sediment

Unconsolidated sediment will collapse if an unstable slope is produced, by direct erosion (eg. stream incision), by removal of the immediate support (eg. stalagmitic floor, rockwall, etc.) or by changes in the mechanical properties of the sediment brought about by weathering, water content, etc. This last set of factors, which does not involve a steepening of the slope but merely a drop in the stability of an existing slope, is often responsible for the initiation of fast mass movement; these factors will therefore be discussed elsewhere (see p.372). For consideration of the mechanics involved in the collapse of unconsolidated sediment, see Statham (1977) and Brunsden (1979).

When only comparatively localised redistribution of sediment is involved, the result will depend very much upon the texture and cohesiveness of that sediment. Coarse, uncohesive

sediments will form generally chaotic deposits (cf. rock collapse, supra), although dry sands may give moderately well bedded fans, with high angles near the base and progressively flatter slopes upwards. With more cohesive sediments, a variety of structures may result (cf. Reineck & Singh 1980:90-2). Under drier conditions, relatively undisturbed blocks of sediment may fall or high angle faulting may occur. Under wetter conditions, and especially with more plastic sediments, rotational slips, shallow planar slides and a variety of plastic deformation features will occur. The collapse of at least partially cohesive sediments may be termed free slumping. Care must be taken to differentiate between such features and subsurface slumping, which is not associated with an unstable surface slope but rather with the gradual removal of support from below the deposit (see p.609).

The author has often observed sediment collapse in modern cave situations, especially along stream banks and at sections left behind by old archaeological or palaeontological excavations. However, well characterised fossil occurrences have not yet been reported. Although it may be that such deposits have a low potential for preservation, the possibility of their existence should be borne in mind, especially when one is faced with a high angle unconformity of uncertain origin.

9.3. Slow Mass Movement

9.3.1. General

Slow mass movement is probably the most difficult mass movement category to document because the actual processes involved cannot usually be observed in action in the field.

Since individual processes act so slowly, any observed movement is likely to be the combined result of several processes, some of which (eg. wash; see p.399) may not even be mass movement phenomena. Even in the laboratory, once water, in any form, is introduced to a simulation, theoretically distinct processes are difficult to separate.

Nevertheless, it is necessary to give some sort of working definition. Slow mass movement very rarely exceeds an average speed of 1m/yr, and is usually recorded at rates of below 50cm/yr, although 'surges' of short duration and higher speed may occur. The data collated by Benedict (1976) suggest that, in the great majority of frost creep and gelifluction (infra) situations, average surface velocity is below 15cm/yr. Slow mass movement does not involve large scale failure of sediments, as do the fast mass movement processes discussed below in section 9.4.

All forms of slow mass movement result from the banal fact that gravity tends to cause displacement of material downslope. The simplest theoretical process is defined by Embleton and Whalley: "Creep is any time-dependent deformation under constant stress."(1979:29). However, in practice, the term creep (or continuous creep) is further restricted:

> Generally if applied stress is less than residual strength . . . , creep rate declines through time to zero . . . . At stresses lower than the yield stress but higher than residual strength, deformation approaches a constant rate through time. In this stress range, however, there is the probability of progressive failure to be taken into consideration. That is, through time, strength falls to the residual level due to loss of cohesion and fails rapidly. Thus the duration of continuous creep is limited by the length of time required for strength to fall to residual levels. This itself is stress dependent; the higher the applied stress the shorter the duration of steady-state creep. (Statham 1977:107)

The above statement is accurate but a little difficult to follow. The yield stress is that force which will immediately overcome the strength of the sediment and cause it to fail.

As a sediment creeps its strength drops. If this strength never drops to the level of the applied stress, we may say that the residual strength of the sediment is greater than the applied stress. If, however, strength drops to the same level as the applied stress, the sediment will fail; we may say that applied stress was greater than residual strength. In natural situations, stress is not constant, varying for instance with gradient, overburden, water content, etc. Some forms of movement (solifluction and gelifluction), which might theoretically be included under creep processes, are usually treated separately. At the very low velocities of interest here, the distinction between plastic deformation and flow is academic. However, the term 'flow' is usually used in connection with solifluction and gelifluction. Since the characteristics of viscous flow developed during slow mass movement are generally immature, full discussion will be deferred to section 9.4.

The term seasonal creep, used by many authors, in fact covers processes which do not generally correspond, even approximately, to the above definition. Most of these processes involve movement initiated by volume changes in the sediment. Freeze-thaw movement, or frost heave, comprises expansion perpendicular to the surface slope and resettling with a downslope component. Expansion and contraction, due to moisture changes and temperature changes (volume changes in the sediment particles themselves, not in the pore water) also create movement so that gravity can impart a small downhill component.

Random creep refers to processes through which particles tend to move under gravity into any available void, producing a net downhill movement. Voids may form through the action of biological factors (roots, burrowing animals, etc.) or small scale

geological factors (solution, removal of fines by water, etc.).

Detailed discussion of all these processes is provided by Statham (1977) and Brunsden (1979). Because of the complexity of the contributing factors, the following sections will be concerned, quite unashamedly, with our present need to interpret fossil deposits.

9.3.2. The Concepts of Solifluction and Gelifluction

This author seriously doubts the present usefulness of the concept of 'gelifluction' to the cave sedimentologist. This is because the only common denominator in the usage of this term, or of apparent synonyms, seems to be the requirement of the presence of a frozen substratum. The geomorphologist can see this frozen substratum and can therefore usefully group the mass movement phenomena associated with it. The sedimentologist involved with larger open-air sites can often afford to search for unequivocal proof of the presence of frozen ground at a given period in the past. The cave sedimentologist, usually working with very small exposures, will be very lucky if such proof is available to him (cf. p.267 and also Black 1969). It is clear from the literature that the presence of a deposit that would be referred to the action of 'gelifluction' if a frozen substratum could be proven, does not in itself indicate continuous permafrost, does not in itself indicate discontinuous permafrost, does not in itself indicate sporadic permafrost, and, in the case of certain types of deposit, does not necessarily indicate the presence of any kind of frozen substratum at all. 'Gelifluction' is a term which refers to a range of mass movement phenomena occurring within a given range of environments. The processes and resulting

sediment bodies are not restricted to particular environments. This point is so important that it will be worthwhile illustrating the difficulties with citations from the literature. Although these arguments may appear at first sight to be pure logomachy, this will not be the case if the reader keeps in mind the question: 'What effect will this have on the sediments?'.

First, the terms solifluction and gelifluction must be properly introduced. Gelifluction, as has been said above, is only operative if some form of frozen substratum is present. This prerequisite, which is banal and faithfully served up in hundreds of textbooks, is expressed by Embleton and King as follows:

> Permanently or seasonally frozen sub-surface layers prevent downward percolation of moisture. The upper layer affected by seasonal thawing (the 'active layer') becomes soaked with water from melting snow, from melting of any temporary segregations of ice within it, or from rainfall. (. . .) The effect of excess water in the active layer is to reduce its shear strength. (1975:97-8)

There is a clear implication that something is going to happen to this active layer, or to components of it, but from this point on, reference must be made to individual usage.

'Solifluction' is properly defined, ". . . the slow flowing from higher to lower ground of masses of waste saturated with water."(Andersson 1906:95). However, few authors use this term correctly; some may even imply processes and conditions that are more restricted than some definitions of gelifluction itself.

Baulig is credited with the introduction of the term gelifluction:

> . . . en régime cryergique, l'ablation des débris se fait surtout par solifluxion - on dirait peut-être plus précisément par géli(soli)fluxion. Celle-ci consiste dans le déplacement lent de matériaux dégelés, saturés d'eau, glissant sur un sous-sol encore gelé . . . . (1956:50-1)

> [Trans.: . . . in a cryergic regime, removal of debris takes place above all by solifluction - to be more precise, one might even say by geli(soli)fluction. This consists of the slow movement of thawed material, saturated with water, sliding on a substratum which is still frozen . . . .]

Here, the active layer slides over the rigid substratum. Stress is laid on the mechanical, or physical, relationship between two materials with different properties, rather than on the role of the substratum in the prevention of loss of water by percolation.

Embleton and King, having reproduced this same extract from Baulig, go on to state: "The advantage of using this term in place of solifluction is that a periglacial régime is definitely implied."(1975:97). Note that it is the usage of the term 'gelifluction' that implies a periglacial context, not the processes or the sedimentary expression themselves. Embleton and King continue: ". . . there is a growing tendency to use gelifluction . . . , with its more precise connotation, to denote flow of thawed material over frozen sub-surface layers."(ibid., p.97). Baulig's 'slide' has now become 'flow', although the physical limit of the substratum is still implied.

> It is unfortunate that gelifluction has been used to denote the combined effect of periglacial slope processes in general and for a particular process (the flow of thawed debris) excluding frost creep. As will be discussed [elsewhere], there are often practical problems of differentiating processes of downslope movement in the field and also in the laboratory. (ibid., p.97)

Embleton and King would prefer to isolate 'gelifluction', as defined by them, from other processes but they admit that this is often difficult in practice. Are they advocating that the term 'gelifluction' should be reserved for cases in which the process can be isolated, or are they suggesting that it should be seen more as a theoretical process? Embleton and King restate their definition: "[Here], the term gelifluction will be used sensu stricto to signify the flow of water-soaked debris over

permanently or seasonally frozen sub-surface layers, akin to mud flow . . . ."(ibid., p.97). The relationship with 'mud flow' is not specified at this point. No definition of 'mud flow' is offered by Embleton and King but, from discussion (ibid., p.99) of the observations of other authors, it would appear that "very rapid" movement is included. Furthermore: "Such events [mud flows] are not, of course, peculiar to the periglacial environment, and are common in alpine regions of steep slopes, high precipitation and snow-melt."(ibid., p.99). However: "The distinction between mudflow [sic] and gelifluction is arbitrary; it is best to reserve the term gelifluction for superficial movements over a frozen sub-surface."(ibid., p.99). The meaning of this statement is not clear. Either synonymy between 'gelifluction' and 'mudflow', as long as it is over a frozen substratum, is implied, or the first phrase means that the two processes overlap and not that the two terms are interchangeable, even given a frozen substratum. In the latter case, the second phrase must contain some criterion for differentiation between the two. The examples of "mudflow" quoted by Embleton and King refer only to superficial movements and at least one instance occurred in an area of permafrost. The word "over" may have a very literal meaning, i.e. 'directly over'. This would allow 'mudflow' to occur in areas of permafrost, but without there being direct contact between the 'mudflow' and any frozen sediment. Alternatively, it may be the word "sub-surface" that is the key. This would require 'mudflow' to come adrift from its source sediment and flow onto some pre-existing surface, whether frozen or not. In any case, Embleton and King would appear to allow 'gelifluction' at speeds far in excess of the 1m/yr limit for slow mass movement suggested by the present author on the grounds of many published observations.

Whatever the exact opinions of Embleton and King on this matter may be, Brunsden (1979) seems to be in agreement, since his section on gelifluction is largely an unattributed quote of the earlier work.

Statham does not use the actual term 'gelifluction'. However, in his chapter on slow mass movement, he states:

> Creep rates increase towards periglacial areas (permafrost zones), where the process is frequently referred to as solifluction. There is no sharp distinction between solifluction and seasonal creep rates, and considerable overlap occurs between the fastest seasonal-creep and the slowest solifluction rates. (1977:112)

Here, Statham would appear to be defining solifluction as 'fast creep operative in permafrost zones'.

> Mechanisms for solifluction include heave due to freeze-thaw and also flow along zones of thawed-out ice lenses in the soil. An alternative to the interpretation of solifluction as a creep mechanism was raised by Chandler (1972), [who] was able to explain movement within the lobes [in periglacial Spitzbergen] primarily in terms of shallow sliding. (ibid., p.112-3)

Statham is using the term 'solifluction', despite its etymological connotation of flow, to cover at least three separate processes. Note that "flow along zones of thawed-out ice lenses" is not quite the same thing as "flow of water-soaked debris over permanently or seasonally frozen sub-surface layers"(Embleton & King, quoted above).

French completes the possible combinations of these definitions: "When operating together the two processes of gelifluction and frost creep constitute the movement which is generally termed solifluction in the modern sense."(1976:135).

Further examples of the difficulties of working with the concepts of gelifluction and solifluction could easily be cited but the situation would only become even more complicated. No major criticism of the authors quoted above is intended. They are interested in explaining phenomena demonstrably associated with

given environments. We are interested in sediment bodies from which we hope to infer processes, and eventually environments. As the understanding of contemporary processes improves, so we will be able to make inferences on a sounder basis.

9.3.3. Sedimentary Features Resulting from Slow Mass Movement

It has been said above that the author sees no adequate way to link the effects of slow mass movement with the normal classification of slow mass movement processes as used by physical geographers. This is a most unsatisfactory situation, since modern observation must form the basis of our understanding of ancient deposits. Nevertheless, the following discussion is offered, incomplete as it may be.

We may infer that small scale lateral movement must occur in scree deposits from the observation that even thick deposits, with little fine matrix, often lie significantly below the theoretical angle of repose. One way in which this might happen has already been noted, namely the sliding of particles over snow.

The movement of individual stones is sometimes called rock particle creep.

> There are very few field measurements by which we can judge the efficacy of rock particle creep in sediment transport but enough exist to show us that rates are low. If we [consider a] 30° slope of 100 metres' length, a boulder being shifted by insolation creep [thermal expansion and contraction] would take a minimum of 47,600 years to reach the bottom, on measured rates. Similar calculations of velocity for the freeze-thaw mechanism of a rock creep reported by Schumm (1967) yield a minimum travel time of about 1450 years, and for freeze-thaw in fine-grained screes a minimum travel time of 250 years is about right. (Statham 1977:103-4)

Rate of creep would therefore seem to be inversely proportional to particle size. This would appear reasonable, because smaller particles require less energy to move them.

Also, finer deposits tend to contain more matrix and have greater water retention capacity, a feature which facilitates frost heave. There is positive feedback here if the rock particles are susceptible to cryoclasticism; the particles are comminuted and the water retention capacity is further increased. If the scree becomes sufficiently impervious to allow significant flow of water over the slope surface at certain periods of high water availability (rainstorm, rapid melt, etc.), stones and fines may be directly transported (see wash processes, p.399). Melting snow might also settle into scree and produce a short period of water flow at or near the surface (see grèzes litées, p.310). The movement of scree would therefore seem to be largely a function of its particle size, including that of any fine matrix. The various processes of production and movement in scree often lead to a slope zonation: (1) feeder cliff; (2) high angle primary scree; (3) lower angle zone of mass movement (creep, solifluction, etc.); (4) basal concave section with wash deposits (Jahn 1960). None of the processes noted above, save sliding over snow, would appear to be capable of moving clean, coarser scree to any significant extent. With thick material lying at angles appreciably below the theoretical angle of repose, the possibility that an original fine matrix has been later removed should be reconsidered (see p.439).

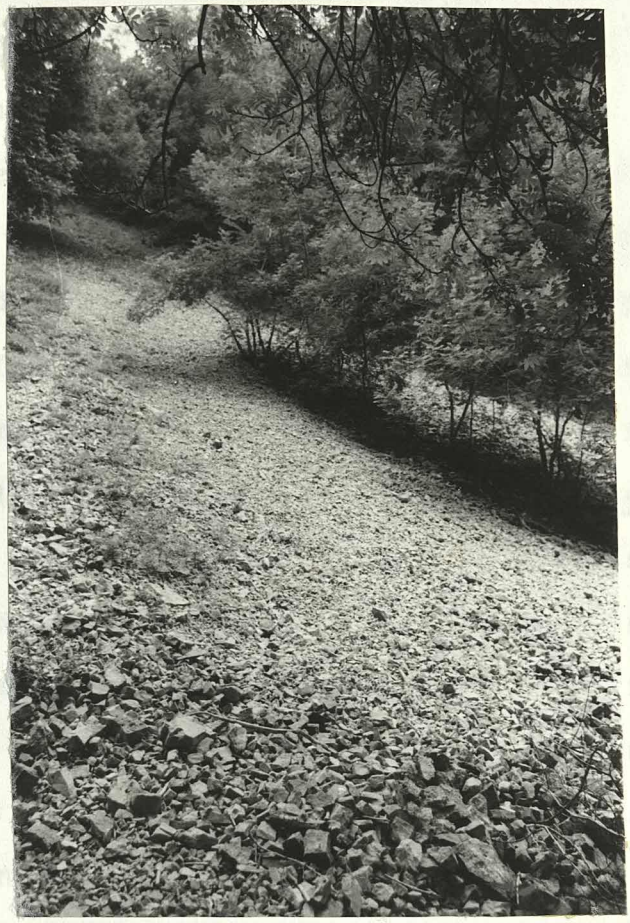
If the interstices in clean scree were filled with ice, something approximating to a rock glacier would form. However, the scale of known rock glaciers is a problem. If we assume a very thick scree (but a very thin rock glacier) of ten metres and we use parameters for bulk density, viscosity, etc., derived from the study of rock glaciers (cf. Embleton & King 1975:122), a point on the surface of a scree sloping at 30° would travel 100m

in about 7000 years. This is much slower than the frost heave figures given by Statham. Significant movement of this type could therefore only be associated with huge deposits, probably recognisable from their scale alone (cf. S.E. White 1976a).

A more realistic process might be movement of the scree by deformation of underlying deposits. This should be reasonably easy to spot, since there would not only be deformation structures in the underlying deposits, but the scree itself would deform. If movement was intermittent and rapid, the scree should show many collapse features, especially near any rock face; the underlying deposits might contain shear planes. If movement was much slower but more continuous, one might expect curved or undulating bedding to develop in the scree, perhaps with areas of very high angle imbrication. Such movement would not, of course, be restricted to clean scree (see Rhinoceros Hole, p.975). Very similar deformation might be associated with the growth and decay of ice in the underlying deposits - a sort of basal 'gelifluction'. Identification of such a process would depend, amongst other things, upon an argument that the overlying disrupted scree did not represent sufficient overburden to produce movement without the aid of ice, an argument which would be very difficult to quantify. Despite these possibilities of large scale, deep seated movement, it should be noted that, at least in the short term, observable movement in scree is concentrated in the surface zone, rarely beyond a depth of a metre (Brunner & Scheidegger 1974).

Sedimentary features definitely associated with the movement of modern scree have very rarely been isolated. Many of the structures noted in 'primary' scree (see p.320) may be enhanced or even created by slow mass movement. Elongate stone chutes are often observed beyond the scree talus itself, but

normally only if slope angles remain high. Loughran and Loughran (1979) report chutes on limestone from N. Yorkshire at 30° to 37°. They were developed where the joint density in the source limestone was greatest, suggesting that they are associated with zones of high initial scree production. Material of mean diameters of between 2.22 and 4.50cm was being moved, but there was no downslope size sorting or change in sphericity, both of these parameters being similar to those of the source scree. In the upper part of chutes coarser stones formed 'levées' at the margins, with finer stones and a little silt and colloids (5-11%) in the centre. Further down the chutes sorting perpendicular to the long axis of the chute was still apparent, although no levées occurred. Imbrication was common, with the lines of imbricated material parallel to the slope. Movement was greatest in the winter months (Nov.-Jan.); it was intermittent and appeared to be associated with snow melt and frost. Average annual rates of movement are not available but, between 14th. July 1976 and 27th. April 1977, monitored stones moved 39.5-73.6cm. One of the problems involved in interpreting highly organised deposits like these lies in the fact that, although slow mass movement is demonstrably occurring now, there is no way of telling how much of the organisation is fossil and perhaps the result of much faster movement in the past. Compare, for example, the patterning seen in stone chutes and that of debris flows, discussed in section 9.4. below. Only very long term studies can provide solutions to this problem, preferably including a 'zero setting', in this case a systematic destruction of the organisation of an existing chute. The subject of block fields and block streams is a related topic; the reader is referred to S.E. White (1976a) for a review concerning such deposits.



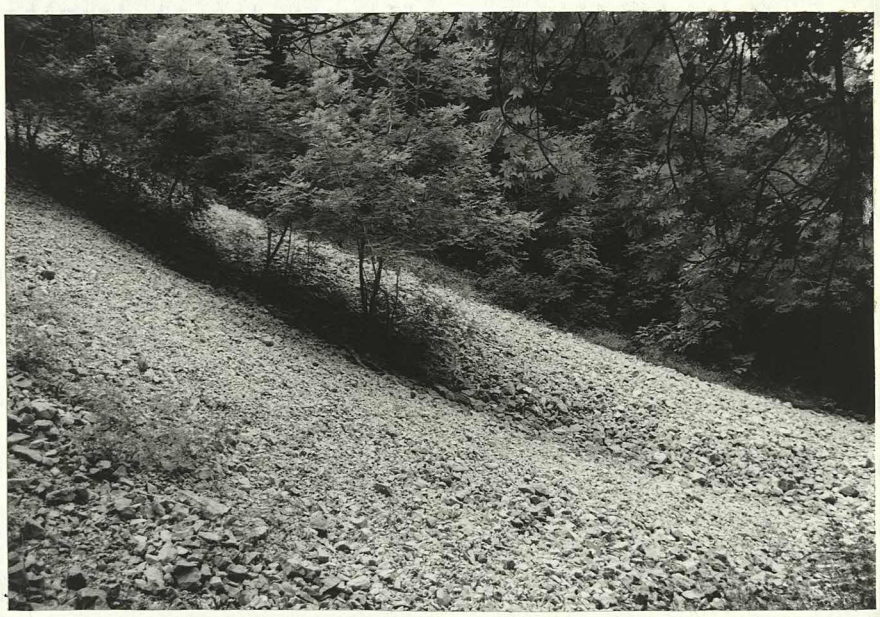


Fig.23 Ebbor Scree

Scree slope within the old quarry at the west end of Ebbor Gorge. Note the relatively straight long profile, component lobes, size sorting (down and across lobes) and weak imbrication.

We now come to slow mass movement in sediments with more abundant fine matrix. This brings us face to face with all the problems evoked in section 9.3.2. above.

The choice of 1m/yr as an upper speed limit is based upon reports of modern sediment movement. When sediment moves faster than this, it tends to move at very much greater speeds; there are few reports of movement between 1m/yr and about 1m/hr (with the exception of mudslides; see p.367). However, it may be extremely difficult to differentiate between the deposits of faster 'slow mass movement' and of slower 'fast mass movement'. Logically, there must be differences and we must try to recognise them, since we must be able to provide a rough estimate of the speed in order to reconstruct the events at the time of movement. Slow mass movement tends to be a continuous or seasonally recurrent phenomena, repeatedly affecting the same sediment. Fast mass movement is usually a single event resulting from failure; the sediment is not normally remobilised by the same process once it has ceased to move, although there may be a rapid series of events, each with its own initial sediment input. This difference is of the greatest importance to environmental and taphonomic studies.

The first priority is a detailed discussion of why material moves. Despite the confusion over the exact meaning of the term 'gelifluction', it is clear that the presence of snow and ice greatly accelerates the development of slow mass movement deposits. These factors will therefore be included in the discussion but it should be remembered that subzero temperatures are by no means restricted to periglacial regions.

During freezing, particles are moved by the expansion occasioned by the transformation of water to ice and, more

importantly in many sediments, by ice segregation. Movement is roughly perpendicular to the surface. Different sediments react in different ways, a characteristic known as frost susceptibility (see p.260). We may recall here that, as a generalisation, finer sediments (clays and silts) are the most frost susceptible and show the most ice segregation; sediments containing under c.10% of material below 20 microns are hardly frost susceptible at all. Freezing also disrupts the colloidal structure of fine material (Tricart 1969). Thus, not only is actual movement caused by freezing, but the structure (cohesion) of fine sediments is also disrupted and water content is raised (ice segregation). These last two factors become important when melting sets in.

During the thaw an already weakened (less cohesive) material will be subject to approximately vertical settling under gravity (loss of volume as segregated ice melts). So far the process might be described as frost heave. Displacement at right angles to the surface during freezing but in a more vertical direction during thawing imparts a small net downslope movement: frost creep. What happens next is dependent upon the quantity of water present in the sediment, upon dynamic properties of the sediment and upon slope.

In the following, the term 'solifluction' will be used to cover all the processes, including flow and frost creep, that together result in recognisable large scale sediment bodies. This is merely an expedient and more exact processes will be indicated wherever possible. 'Gelifluction' will be considered as a subdivision of solifluction, operative when there is a significant ice content in the lower part of the sediments. Direct quotes cannot be changed to fit these definitions but a note will be added if clarification is necessary.

Water content may be considered to be the balance between input (melting ground ice, melting snow, rainfall, etc.) and the possibility of rapid loss. If water can escape at a rate approximating to input, there will be little further movement. If water cannot so escape, the active water content of a sediment will rise. Water loss may be inhibited by a substratum of low permeability (which may be frozen or naturally impermeable), by an impermeable superstratum (eg. a snow bank, superficial freezing, etc.) and by water retention properties of the sediment itself. Stress in the sediment is a function of slope, the bulk density of the sediment/water mix and any external pressure (eg. from an overlying snow bank or superficial freezing). The resulting strain is dependent upon dynamic properties of the sediment/water mix, principally concerned with whether the material behaves more like a plastic, a (Newtonian) fluid or a brittle solid. Plastics deform indefinitely once a threshold stress is applied. Strain in fluids is a linear function of applied stress. Brittle solids fail with little or no deformation at a threshold stress. As water content increases, stress increases (rising bulk density) and the dynamic properties change. If the material behaves most like a brittle solid, failure will eventually occur and the result will be a slide or slip, often followed by fast mass movement. Solifluction occurs in materials showing pseudo-plastic behaviour, that is, behaviour somewhere between true plastic and true fluid. These materials creep ("any time dependent deformation under constant stress", supra). Under constant stress, once creep has started strain will continue to rise until the material eventually fails, again producing sliding and usually rapid flow. Therefore, by our definition, slow mass movement involving true creep can only be significant if something happens

to prevent the otherwise inevitable failure. There are three factors which can achieve this. Stress may decrease (movement to a lower slope angle, drop in bulk density through water loss, etc.). The dynamic properties of the mix may change (loss of water, disintegration of sand-sized aggregates of clay, etc.). An external factor may prevent movement (a superficial turf layer, the next freezing episode, etc.). Solifluction is therefore a delicate balance of factors; if the balance is disturbed, failure and fast mass movement may result.

The sediments that are likely to move in this way are of many types. Under optimal external conditions just about any sediment can creep, since no material is perfectly brittle, fluid or plastic. However, extensive sediment movement is more likely to occur in more plastic sediments. This property is expressed in terms of the plastic limit (W_P), the liquid limit (W_L) and the plasticity index (I_P). These factors are theoretically defined as follows: W_P is the water content below which the sediment no longer behaves as a plastic; W_L is the water content above which the sediment behaves as a viscous fluid (or 'liquid'); $I_P = W_L - W_P$. In practice, the limits for a given sediment are assessed by <u>arbitrary</u> tests (cf. Bowles 1979; Linell & Tedrow 1981), involving direct observation of how a sediment actually behaves in a laboratory. The limits cannot be accurately inferred from other simple parameters, such as particle size distribution, since a very large range of sediment parameters is in fact involved. What is more, the laboratory tests can only be carried out on the fine matrix of a sediment (defined differently in different disciplines and in different countries, but in this connection usually material under c.0.4mm), since realistic results could only be achieved for whole sediments if enormous

samples were used and new arbitrary tests and equipment devised. In theory, sediments which have a low W_P and a high W_L (i.e. sediments with a high I_P) should creep best. However, there is quite a high degree of positive correlation between W_P and W_L. Since the test for W_L (involving special equipment) is more objective than that for W_P (involving considerable manual skill), only the former limit is usually quoted. This limit, of course, represents the point at which slow mass movement must change to fast mass movement. Many solifluction deposits have low liquid limits, and thus low plastic limits.

Great care should be taken when using the concept of liquid and plastic limits. Many fossil sediments, especially calcareous ones, will have developed properties very different from those that governed their original movement. Very slight changes in carbonate content and in the exact size and geometric distribution of carbonates may invalidate limit testing in most cave sediments. The present author predicted and found a significant relationship between carbonate distribution resulting from diagenesis and limit values in otherwise homogeneous debris flow deposits at Pontnewydd Cave. Only when the sediment can be realistically assumed to have been free of carbonate and organic matter at the time of emplacement, and when there has been no major alteration (chemical or mechanical) of the sediment, can 'cleaned'(HCl, NaOH, etc.) sediments be used for testing. The cleaning itself can even change the properties of interest if active clay minerals are present.

Only very basic rules of thumb are therefore available for assessing whether an ancient sediment would have been likely to move by solifluction. Direct testing of frost susceptibility may suffer from the same drawbacks as does limit testing. However,

since frost susceptibility is much more dependent upon particle size, useful approximations can be made from modern or theoretical data (see p.260). Susceptibility to creep would appear to be greater if fine materials, especially clay minerals, organic matter and fine silts, are present. Particle interlocking is reduced if the sediment is badly sorted, a factor which may also encourage ice segregation (see p.261). This lack of a precise method of assessing susceptibility to creep in ancient sediments is probably the single most important factor in our present inability to refute claims of solifluction and gelifluction deposits (infra). The influence of slope has frequently been mentioned above and this is, of course, one of the parameters that can often be directly measured in ancient deposits. However, since slope is only one parameter in a very complex system, there is little direct relationship between deposit and slope. All other things being equal, movement may be faster on steeper slopes. Solifluction has been observed on slopes ranging from 35° to 1° (Benedict 1976; Washburn 1979), that is, it has been observed at almost every natural slope angle. The majority of cases probably fall between 25° and 5°; at steeper angles deposits are unstable and often fail and at lower angles movement is increasingly dependent upon gliding over thawing ice (Washburn 1979).

Keeping the above discussion of processes in mind, we will now turn to a description of the deposits themselves.

Gelifluction deposits are closely related to local rock type since it is rare for material to have travelled farther than 1 or 2 km. by the processes described. Thus, unlike glacial drift, erratics are rare (unless the source of the material is drift itself). Angularity of the fragments is characteristic since they have not travelled far, though some reduction in angularity with increasing distance from the source has been occasionally noted (for instance, T.T.Paterson, 1951, p.11). The thickness of the deposits is extremely variable. Depressions and valley bottoms may collect considerable infills, sometimes several tens

> of metres thick. The longer axes of stones in the deposit lie in the direction of movement (compare glacial drift) until the point is reached where movement is arrested, as at the edge of a lobe or terrace. Then the blocks are suddenly turned at right-angles to the previous movement (G.Lundqvist, 1949). Reorientation at the border is facilitated by the fact that blocks never reach the front of the moving mass at exactly 90° to it. Obstacles within the moving mass also cause local reorientation of stones. Flow orientation is less persistent and regular in the lower layers of such deposits, unless the latter are derived from glacial drift, when orientation in the lower layers may express the last direction of ice motion. Signs of bedding are usually rare: head deposits [a generic term used for deposits thought to have resulted from solifluction] are typically massive and unsorted. The origin of occasional indistinct bedding is not clear, but it may represent accumulation in different episodes and may be related to changing climatic conditions.
>
> Contorted structures or 'involutions' are often typical of those deposits where bedding can be distinguished. Involutions associated with ground-ice formation were considered [elsewhere; see p.272 in the present text]. Downslope movement can itself produce deformation or drag structures owing to differential lateral movement; K.Bryan (1946) distinguished these as 'plications'. A.Jahn (1956) shows stages in the development of such forms, the final stage being characterized by roll-like or cylindrical masses . . . . (. . .)
>
> Several problems arise in distinguishing head deposits from glacial drift . . . . (. . .)
>
> An example of typical Quaternary head deposits from south-west England is given by D.N.Mottershead (1971). (. . .) All samples analysed fell within the very poorly sorted or extremely poorly sorted categories, suggested short-distance transport en masse, but with increasing distance from possible source areas [blockfields and tors], there is a tendency for the head to become finer and less stony as some comminution of the debris occurs and some of the coarser blocks are left behind. All samples fell within the range of frost-heaving materials and were characterised by low liquid limits, such that only 25 per cent of water by weight of sediment would suffice to liquify it and cause flow. (Embleton & King 1975:106-7)

This description is clear, apart from the last statement. Had water content risen above 25% the sediment would, at least by some definitions, have been liquified and only special circumstances would have prevented fast mass movement. It has already been said above that the present author considers the distinction between plastic deformation and flow to be academic in the context of solifluction. Thus flow, creep, plastic deformation or whatever one cares to call it would have occurred before the liquid limit was reached.

The most important point brought out in the above extract is that many of the characteristics of solifluction deposits are

inherited from the source. Embleton and King note that, apart from local rock types and the regolith derived from them, glacial sediments can act as the source and, in so doing, can impart special characteristics. In fact, any sediment source will have a similar effect. Thus, petrology, angularity, fabric, sorting or bedding in the source sediment may be retained or partially retained in the solifluction deposit. Nevertheless, the great likelihood, particularly on slopes in limestone areas that have been subject to a dominantly erosive regime for long periods, is that sources will be purely local and themselves rather unstructured; in pedological terms, they are both lithosolic and regosolic. A similar situation occurs in the Arctic today:

> Soils on the sloping lands may be considered somewhat 'Regosolic' in character, in that they display few genetic features, but there is [still] a considerable range of conditions associated with solifluction slopes including that of substrate and drainage. (Linell & Tedrow 1981:43,45)

We need to know how far solifluction is capable of altering the source parameters.

In the following we must differentiate, at least in theory, between flow phenomena and frost heave. Given a well bedded source, with layers of different lithology, there is reason to believe that both flow and heave phenomena might at first preserve some traces of the bedding, since the various layers will each have a different response to these processes. However, this differential response must itself cause increasing distortion, by very definition, giving rise to involutions and drag structures. Any major inhomogeneity in the sediment must inhibit a coherent flow pattern, especially if convoluted layers of different viscosity are present. Thus, it may be inferred that a certain amount of homogenisation will be necessary before a regular pattern of movement can be achieved. Distortion will continue with the

component sediment bodies becoming more and more fragmentary and the boundaries less distinct. If the material were not on a slope, frost sorting mechanisms might eventually impose a new structure (see p.273), but this should be largely inhibited by the decreasing resistance to coherent flow produced by the breakdown of the original bedding. However, quite strong surface patterning, such as stripes of coarse material, can certainly survive movement; indeed, there is evidence that the formation of stripes rather than polygons is a function of movement (Washburn 1979). We may say, in complete agreement with Benedict: "Gelifluction encourages textural homogeneity . . . ."(1976:62). Furthermore, French has affirmed: "In general, most investigations indicate . . . that solifluction movement is laminar in nature."(1976:137). It is highly probable that traces of any original bedding will decrease away from the source, although they may never disappear entirely. The author is unaware of any study specifically designed to test this assumption but oblique references in the general literature would seem to provide some support.

Embleton and King note indistinct bedding and suggest a possible stratification of different sediment bodies through time. This could also be due to retention of some bedding features from the source, as we have seen above. A further origin might be segregation of sediment components into 'pseudo-layers' during a phase of movement. This would seem rather unlikely, except in the case of coarse material, which might be sorted towards the surface by frost heave or towards the base, either by its own greater resistance to movement or through 'caterpillar tracking' at the flow margins (infra).

Büdel (1959) has suggested that a distinction can be made between periodic solifluction, remobilised each year, and episodic

solifluction, occurring as more or less isolated events. According to Büdel, the periodic type is characterised as occurring on higher slopes as thinner deposits with laminar fabrics, whilst the episodic type occurs on lower slopes as thicker deposits with contorted structure. These observations were based on fossil occurrences in mid-latitudes and their periodic or episodic nature is therefore inferential. Nevertheless, the patterns of slope, thickness and structure seem logical whatever the timing.

The observation that 'head' may be comminuted as it moves away from the source, coupled with the apparent lack of any decrease in angularity, would suggest that frost shattering might be responsible. Some crushing in coarser sediments may also occur, but chemical processes, such as hydration, should not be ruled out. Certainly, the general literature does not contain references to heavily worn stones unless this feature is obviously inherited from the source. For instance, Benedict (1976) states quite categorically: "Neither frost creep nor gelifluction is capable of causing significant changes in the angularity of stones."(ibid., p.61). This contrasts with the observations of Laville (1975) that geliflucted limestone clasts in shelter sequences have undergone significant mechanical rounding (infra).

Preferential orientation, both of stones and smaller elements (eg. sand grains; cf. Benedict 1970), is repeatedly recorded from solifluction deposits. Benedict writes:

> Where movement is unimpeded, a strong downslope long-axis orientation tends to develop in the upper metre of soil. Secondary oblique orientations, at angles of 30-50° with the direction of movement, are also common, as is a minor transverse orientation. Where movement is obstructed, the transverse orientation becomes increasingly well developed. Tabular particles of all sizes tend to lie with their flat faces parallel to the surface of the deposit or slightly imbricated. (1976:63)

Brochu (1978) has compiled average figures for long axis orientation of stones over many deposits from the Arctic; the most

commonly occurring proportion of stones oriented with slope is c.80%. At the front of a flow, not only does long axis orientation tend to become more transverse, but orientation of the minor (thickness) axis also usually changes. If stones are very common at this point, the secondary orientation tends to show as a sharp dip downslope. However, if the front is rich in fines, and especially if there is a vegetation mat, a significant upslope dip may result. French (1976) claims the exact opposite relationship between stoniness, vegetation and dip, suggesting that the real situation is rather more complex than so far recognised. In any case, a considerable degree of imbrication (upslope dip) may show for some distance behind the actual front.

Benedict (1970) has noted a weak platy structure in specifically gelifluсted sandy loam. He has generalised this observation (1976), giving two probable causes: ice segregation (see p.268) and re-orientation of particles during movement. He also notes (1976) vesicular structures, thought to form as small gas bubbles are released from freezing water. Goldberg (1979a), in one of the very few attempts to examine the microscopic structure of cave sediments, has noted several features apparently associated with solifluction (gelifluction?) at the Pech de l'Azé II (Dordogne). Apart from effects referred to freeze-thaw (see p.269), Goldberg describes clay cutans (argillans) c.90 microns thick, with strong continuous clay orientation, coating mineral grains and other fine aggregates, so as to produce rounded clay "clumps" in the order of 0.5mm diameter; individual units may be bound together by cutans into larger composite aggregates. Goldberg suggests that such features may be produced by "rolling" (presumably meaning continuous rotation within a more fluid matrix), possibly during solifluction; an illuvial origin is said to be

unlikely. Despite Goldberg's reticence, it is clear that there is little correspondence between layers with "clay clumps" and layers identified by Laville (1973) as being gelifluction deposits.

Having discussed some of the structural characteristics of solifluction deposits, we must now examine the overall ('mass') characteristics of movement. We have seen that a combination of heave and flow is involved. These are different processes and they produce different velocity profiles (cf. Embleton & King 1975; Brunsden 1979; Washburn 1979). However, the point of importance here is not the exact profile but the fact that movement is greatest at or near the surface and declines quite rapidly to zero with depth. Figures tabulated by Benedict (1976) suggest that, in the great majority of frost creep and gelifluction situations, movement ceases below a depth of c.80cm. Cases where the velocity decline is instantaneous, occurring at a plane through shear, are very rare and are nearly always associated with peculiar source materials (see p.367). Thus, Washburn explicitly writes:

> To the extent sliding as opposed to flow is important in gelifluction, it seems probable to the present writer that it is primarily on a microscale, along thawing ice lenses (Beskow 1930:624-5 [cf. also Carson & Kirkby 1972:277, cited elsewhere by Washburn]) and by soil settlement into voids left by thawing lenses ([C.] Harris 1977:41-2), and that the overall effect is flow, although there are presumably transitions between gelifluction and other processes such as slumping where sliding is usually obvious. (1979:205-6)

In this context, a fundamental misunderstanding on the part of many quaternary sedimentologists may have arisen, concerning the role of permafrost, segregation ice and seasonally frozen sediment bodies as limiting factors for movement. This misunderstanding is deepened by casual statements by physical geographers, for instance, that made by French: ". . . the permafrost surface acts as a water lubricated slip plane for

movement of the overlying active layer."(1976:134). French is correct, in the sense more carefully expressed by Washburn (supra), but he gives the impression, or at least permits the inference, that the active layer may travel for unlimited distances over some apparently even frozen surface. First, the permafrost table (see p.267) is never a smooth planar surface; it is a highly irregular 'zone' with pockets of unfrozen sediment penetrating into the frozen ground and lenses of pure ice and frozen sediment 'surviving' within the base of the active layer. Second, the frost table or tjäle (see p.266) is seasonally variable; during the thaw it may often retreat downwards faster than the measured rates of solifluction. There just is no 'surface' which would remain sufficiently stable and planar for a large body of saturated material to move any distance across it.

The true nature and scale of solifluction is further confused by unfortunate use of terminology, as was noted in section 9.3.2. above. Consider again the statement by Embleton and King: ". . . it is rare for material to have travelled further than 1 or 2 km. by the processes described."(1975:106). All authorities, including Embleton and King themselves, are in agreement that the average velocity of the normal flow stage in solifluction is in the order of 5cm/yr. For a deposit to travel two kilometers at this rate, it would require a stable system to be maintained over 40,000 years, clearly a preposterous suggestion, and one which Embleton and King of course do not make. Their definition of "gelifluction" includes major surges at very much higher velocities, a factor which may well account for their suggesting such long travel distances. But such movement cannot be included in our definition of solifluction as it would make the distinction between 'slow' and 'fast' mass

movement untenable; it is precisely the speed of movement which we hope to be able to deduce from the sediments. Another reason for observing soliflucted material far from its primary source might be separate periods of solifluction occurring intermittently over very long time scales. However, this would not result in structured sediment bodies that have themselves moved over such distances; such bodies will have been created and destroyed many times during the overall movement. As sedimentologists, we must first concern ourselves with the parameters of movement clearly evidenced in a given deposit and only second with the recognition of the ultimate source of the component sediment. Therefore, let it be clearly understood that the processes being discussed here result in a total displacement, over many centuries, of a few metres or a few tens of metres, and no more.

The sediment bodies formed by solifluction take various shapes. They have been described as sheets, benches (terraces) and lobes. 'Gelifluction streams' (Washburn 1979) are markedly linear (downslope) features; the example figured by Washburn (1979:210) is almost certainly a fast mass movement phenomenon and the present author doubts the usefulness of this category within a classification of true solifluction forms. The exact shape of a solifluction deposit is a complex function of such features as topography, dominant slope, sediment type, vegetation, and solar aspect. Even such apparently unlikely factors as wind direction can be involved (cf. the location of snow drifts which first provide insulation and then localised melt water). It would be totally unrealistic to believe that all, or even a majority, of these parameters could be deduced from fossil deposits. The important point here is to recognise that the relationship between roughly planar surfaces ('treads') and sharp changes of

gradient ('steps') at sheet, bench or lobe fronts is a function of the flow pattern.

One of the ways of demonstrating the velocity profile of a solifluction deposit is to place a vertical column of separate markers in the sediment and wait for the line to distort. Markers nearer the surface will move downslope faster, and thus more, than markers lower down; the lowest markers will not move at all. One can imagine an infinite area of slope moving in this way. But the system remains stable only if the velocity profile remains areally constant (i.e. if all conditions for movement are areally constant). Once part of the slope mantle slows or stops it will be overridden by material from further up the slope - this is the origin of the front morphology. It has often been observed that surface stones may move faster than the rest of the sediment and that they may therefore accumulate at the deposit front to form stone-banked terraces. Ploughing blocks are stones that ruck up the sediment or turf ahead of them and leave shallow depressions behind them, providing a graphic illustration of their relatively faster movement. Benedict (1976) suggests from observational data that stones may be expected to move in this way only when frost creep rather than flow is dominant. On theoretical grounds (see p.377), the present author would not dismiss the capabilities of flow so lightly.

From the above discussion, we may expect two quite different types of lower boundary to occur. In the area where flow is approximately constant, the lower boundary must be gradational, between an undisturbed (during this particular phase of movement) substratum and an increasingly dislocated overburden as one approaches the surface. Once the flow pattern changes and overriding occurs, the lower boundary of the overriding mass

becomes abrupt. Overriding may occur by 'caterpillar tracking', with the faster moving surface material rolling down and under the advancing front. Sliding may also occur at this stage, with the overriding sediment moving out en masse over the downslope surface (cf. Benedict 1976:64); this is possible because it is a matter of a pre-existing surface that will be more or less consolidated. In most open-air situations, there will be some vegetation cover. Thus, Benedict writes:

> Virtually all lobes and terraces contain at least one buried organic layer, representing soil humus overrun by the deposit during its downslope advance . . . . Examples have been reported from periglacial environments throughout the world. (1976:64)

Such organic material may, of course, be much rarer in shelter and cave situations. However, the importance of this observation here is that buried organics can be dated. The estimates of velocity given so far apply to the maximum surface rate and they are therefore strictly relevant only upslope of any overriding mass. Once a mass starts to override the downslope surface, we need to know how fast it can travel as a body, how fast the front can advance. Radiocarbon dates on buried organics show that this velocity is at least an order of magnitude slower than the maximum surface velocity. Benedict (1976) gives the figure of 3.5mm/yr as the upper speed limit for overriding masses calculated from radiocarbon dates by a variety of authors in several parts of the world.

Finally, we need to know whether solifluction can be in any way erosive. In that it results in net displacement of material, the answer must be that it can. But can a solifluction mass which is overriding a stable surface cut into the underlying deposit? In other words, when we observe a solifluction deposit that has an abrupt lower boundary, can this boundary be erosional

over any significant distance? The answer to this question must be 'no'. There is simply no available mechanism. Solifluction is erosive by a gradual extension of dynamic characteristics into new sediments; it cannot produce sharp erosive boundaries as can fast mass movement or running water. The only possible form of 'abrupt' erosion would seem to be entrainment of stones and slight planation of surface irregularities by the solifluction front - a minor 'bulldozer' effect.

The concept of solifluction has been variably employed by cave sedimentologists.

Lais (1941) did not recognise solifluction or any other form of more fluid mass movement in caves.

Vértes made the following comments:

> Mit der Solifluktion als Höhlensedimente transportierender geologischen Kraft befasste sich in erster Linie J.KEREKES [1941, 1951], der auf Grund seiner Beobachtungen an mehreren ungarländischen und ausländischen Höhlen feststellt, dass der vom Wind eingeführte Löss zwar in einzelnen Fällen eine gewisse Rolle in der Entstehung der Sedimente hat, doch gelangt das Material der Ausfüllungen in anderen Fällen - und diese bilden die Mehrzahl - von aussen her, durch Solifluktion in die Höhlen. In diesen Fällen wäre auch der Kalkschutt in der Ausfüllung kein Frostbruchprodukt der Höhlendecke, sondern allochton. Als Beweis dafür hebt er hervor, dass die Felswände einzelner Höhlen intakte, uralte Korrosionsformen bewahren, dass die Ausfüllung sich in solchen Höhlen leicht (in einem Wilkel von 2-5°) nach dem Inneren hin senkt, und dass in der Regel auch Kalkschutt fremden Ursprunges im Sediment zu beobachten ist, usw. Seiner Ansicht nach hat sich der Schutt in solchen Fällen durch Kamine oder in der Richtung des Hanges durch den Eingang in die Höhle ergossen. - Unseren Beobachtungen nach steht es ohne Zweifel fest, dass sich Höhlenausfüllungen auch auf diese Weise bilden, doch nur in seltenen Fällen. Dazu, das die Ausfüllung durch Solifluktion in die Höhle gelange, wäre ein öfteres Gefrieren und Auftauen der Sedimente nötig. Dies erfolgt jedoch bereits in geringer Entfernung vom Höhleneingang einwärts bei der ausgeglichenen Temperatur der Höhle nicht mehr. Am Schutt, der durch Solifluktion in die Höhle gelangte, müssten gewisse Spuren der Verschleppung zu bemerken sein; der Schutt müsste schichtenweise gelagert sein, im allgemeinen müsste sich eine Strukturbildung zeigen. Solche Erscheinungen sind jedoch in ganz Europa sehr selten beobachtet worden* [a note; see below] und wurden in ihrer typischen Entwicklung, unseres Wissens, in ungarischen Höhlen bisher überhaupt nicht angetroffen. (1959:15)
>
> * Krioturbationserscheinungen werden von PROŠEK in den Schichten der Dzeravá skala erwähnt [1951]: J.SCHANDLER fand in der Eislueg-Höhle

(1 450 m.ü.d.M.) polygonalen Tundrenboden [1931]. H.ALIMEN beschreibt in einer Felsnische in der Charente Tundrenerscheinungen [1950]. J.KEREKES beschreibt in der Homoródalmáser-Höhle mit Brushit ausgefüllte Frostkeile. (ibid., p.164)

[Trans. Translating from Vértes (1959) is particularly difficult as the German is extremely complex and there appear to be many typographic, and possibly even grammatical, errors. No attempt has been made to correct apparent errors in the original for fear of compounding the possibilities for misunderstanding. Vértes' native language was not, of course, German.

Solifluction as a transport process for cave sediments has been especially studied by J.KEREKES (1941,1951). From his observations in several caves in Hungary and elsewhere, he established that, in individual cases, loess introduced directly by wind action did indeed play some role in the formation of the sediments. However, in the majority of cases, the material of the fill was derived from outside and had been introduced into the cave by solifluction. In these cases, the limestone scree in the fill was also allochthonous and not derived by frost action from the cave roof. Amongst his proofs, he pointed out that the rock walls of each cave preserved very old corrosion forms intact; that the fill in such caves dipped slightly (at an angle of 2-5°) towards the interior; and that, as a rule, limestone scree of exotic origin was also to be noticed in the sediments. In his opinion, the scree was discharged into the cave through chimneys or down slopes and through the cave entrance. Our own observations on this matter remain unchanged: cave fills do originate in this way, but only in a few cases. Furthermore, if solifluction is to penetrate the cave, rather frequent freezing and thawing of the sediment is necessary. However, the decline in temperature variation even a short distance inwards from the cave entrance will prevent this. As for scree introduced by solifluction, certain indications of movement must be observable: the scree must occur within layers [be bedded] and in general must show the development of structure [fabric]. Nevertheless, such phenomena have very seldom been observed in the whole of Europe* and, to our knowledge, have not yet been encountered in really typical forms in the Hungarian caves.

* Cryoturbation phenomena were noted by PROŠEK (1951) from the layers of Dzeravá Cave. J.SCHANDLER (1931) found tundra polygons in the Eislueg Cave (1400m O.D.). H.ALIMEN (1950) described tundra phenomena [polygons] in a rock shelter in the Charente. J.KEREKES described ice-wedge casts, filled with brushite [$CaHPO_4.2H_2O$], in the Homoródalmáser Cave.]

It is clear that Vértes saw solifluction as being linked with frozen ground, though not necessarily with permafrost. Because of this, he came to what would seem to the present author to be a most reasonable conclusion, namely that solifluction is likely to be a rare event in caves. He shrewdly pointed out the need to observe structures or fabric in order to identify such events. But he failed to explain Kerekes' remarkable observations (cf. the arguments used by the present author at Pontnewydd Cave),

and in so doing failed to grasp the logical consequences of his own argument: perhaps mass movement, whether slow or fast, can occur without frequent freezing cycles. The following extracts show how Vértes, a researcher of considerable integrity, was never happy with this situation.

> Um auf die morphologischen Eigentümlicheiten des Schuttes zurückzukehren, finden wir, dass es ausser dem scharfkantigen und dem korrodierten Schutt in einigen seltenen Fällen (im Fälle der untersuchten Höhlen bei der Felsnische von Tarkő, in der Lökvölgyer Höhle und in der grauen Schicht der Bivak-Höhle) auch noch abgerundeten, erodierten, fast kieselartigen Schutt geben kann.
> Wir kennen keine geologischen Kräfte, die innerhalb der Höhle eine so weitläufige und intensive Erosion produzieren könnten, selbst dann nicht, wenn wir die Möglichkeit des "trockenen Polierens" (charriage à sec) [Koby 1938]* im Falle einzelner Knochenfragmente mit in Betracht nehmen [Bächler 1940]. Es muss angenommen werden, dass der kieselartig abgenützte Kalkschutt allochtonen Ursprunges ist, und wahrscheinlich durch Solifluktion in die Höhle gelangte. Auch in diesem Falle hatte er sich mit Frostbruchschutt zu vermengen; der Boden-Transport durch Krioturbation fand ja eben in solchen Zeiten statt, als auch der Frostbruch am Stärksten an der Arbeit war. Wahrhaftig finden wir neben den abgerundeten Kalkieselsteinen immer varhältnismässig viel scharfkantiges Schuttmaterial mit unverwitterten Oberfläche.
> (ibid., p.32)

> [Trans.: Returning to the morphological characteristics of the scree, we find that, in addition to angular scree and corroded scree, truly rounded, eroded, almost pebble-shaped scree can also occur in some rare instances (cf. the Tarkő Rockshelter, the Lökvölgyer Cave and the grey layer of the Bivak Cave, as discussed below).
> We know of no geological process which can produce such extensive and profound erosion inside a cave, even allowing for the potential of 'dry polishing' (charriage à sec) [Koby 1938]* as observed in the case of individual bone fragments [Bächler 1940]. It must be assumed that this limestone scree, worn to pebble-like shapes, is allochthonous and that it was probably deposited in the cave by solifluction. Furthermore, in this case mixing with frost-fractured [autochthonous] scree had to occur; sediment transport by cryoturbation in fact took place during just such periods as these and cryoclasticism was also at its strongest. Indeed, besides the truly rounded limestone gravel we always find comparatively large quantities of angular, unweathered scree.]

> LÖKVÖLGYER HÖHLE
> (. . .) Eine Untersuchung des Materials ist jedoch trotzdem von Wichtigkeit, da sich KEREKES mit dem Material dieser Höhle befasste und es als infolge von Solifluktion in die Höhle gelangtes Material betrachtet [Kerekes 1938]. Er meint, dass es "Schlamm", d.h. ". . . Lehm, aus auf periglaziales Gebeit fallendem Staub gebildet" sei, das sich ". . . infolge klimatischer Gründe (feuchteres Wetter) sich nicht zum Löss umbilden konnte"[Kerekes 1938:119]. Obwohl das Endresultat der petrographischen Untersuchungen auf ein Interstadial als Entstehungszeit hinweist, welches vielleicht kühler als im allgemeinen war, nehmen wir auf Grund der Kornfraktion und der Kalkschuttuntersuchungen die genetische Hypothese von KEREKES an. Es soll jedoch hinzugefügt werden, dass obwohl die

* Vértes's footnote gives a muddled reference, in reality involving two publications covering the same subject: Koby 1938 and 1941.

ursprüngliche Herkunft dieser Schicht - bzw. verallgemeinernd: dieser Art von Schichten - spurenweise nachweisen ist, wird ihr Charakter dennoch von der Zeit der Umlagerung bestimmt, also von jener Periode, in deren Verlauf das betreffende Material die Oberfläche des Höhlensediments bildete. Die durch Erosion abgenützten Kalksteine deuten die Solifluktion an: der Korrosionsindex übersteigt daher die Toleranz eines jeglichen bisher bekannten klimatischen Abschnittes . . . . (ibid., pp.108-9)

[Trans.: LÖKVÖLGYER CAVE
However, an examination of the material is of importance, since KEREKES has dealt with the deposits of this cave and he considered them to be due to solifluction. He believed the material to be "silt", that is ". . . loam, formed from dust falling onto periglacial terrain", and that it was ". . . the result of wetter climatic conditions that would prevent its transformation into loess [structural sense]"[Kerekes 1938:119]. Although the final results of the petrographic analyses point to an interstadial which was perhaps cooler than usual, we accept the genetic hypothesis of KEREKES on the grounds of the grain size distribution and the limestone scree analysis. However, it must be added that, although rare traces of the origin of this layer - and of layers of this sort in general - may still be identified, its character was nevertheless imparted at the time of translocation, the period when the material in question formed the surface [top unit] of the cave sediments. Limestone worn by erosion indicates solifluction; the corrosion [rounding] index therefore lies outside the range of all climatic phases as yet recognised.]

TARKŐ-FELSNISCHE
(. . .) Den petrographischen Untersuchungen zufolge . . . besteht die obere "lössähnliche" Schicht z.T. aus einem Material, das durch Solifluktion umgelagert wurde. In der Entstehung dieses Materials hatte auch der Löss eine Rolle. Das Klima zur Zeit der Umlagerung hat den Charakter des Ende eines Interstadials, oder noch eher eines ozeanischen Abschnittes. Die mittlere gelbe Schicht gelangte nicht durch Solifluktion in der Höhle: der in ihr enthaltene Kalkschutt ist von der üblichen Art. Die Untersuchungen der Schicht deuten entschieden die Krioturbationsverhältnisse an. Die untere braune Schicht gelangte wieder infolge von Solifluktion in die Höhle: sie führt entschieden interglaziale Charakterzüge.
Die Rolle, die die Solifluktion in der Genese der Höhlenausfüllungen hat, kann vorderhand nicht näher erörtert werden. Es kann angenommen werden, dass die Solifluktion ihre Wirkung im Interstadial ebenso, oder annähernd ebenso ausübte, wie in den ozeanischen Abschnitten, die die Vereisungen einleiten. Es ist offensichtlich von lokalen Gegebenheiten bedingt, wann die Solifluktion Material in die Höhlen lieferte. Es kann ja nicht einmal behauptet werden, dass diese Erscheinung in Felsnischen öfters auftritt, und auch das Vorkommen dieses Ereignisses kann weder an eine Region, noch an andere Bedingungen gebunden werden. (ibid., p.110)

[Trans.: TARKŐ ROCKSHELTER
From the petrographic analyses . . . it can be seen that the upper loess-like layer is in part formed from material brought in by solifluction. Loess has also played a part in the formation of this material. The climate at the time of transportation shows the characteristics associated with the end of an interstadial or, even more likely, of an oceanic period. The middle yellow layer was not deposited in the cave by solifluction; the limestone component is there of the usual [angular] kind. Examination of this layer revealed clear indications of cryoturbation [unspecified]. The lower brown layer again arrived in the

cave through solifluction; it shows clear interglacial [sic] characteristics. [This last layer contains "stark abgerundete[n], kieselförmige[n] Kalkschutt" (strongly rounded, pebble-shaper limestone scree); ibid., p.110.]
The role played by solifluction in the genesis of cave fills cannot be discussed further for the present. It is probable that solifluction continued in the same way, or approximately in the same way, during the interstadials as it did in the oceanic periods which mark the onset of ice build-up [glacial stages but not necessarily the glaciation of the actual areas under consideration]. Local conditions obviously controlled when solifluction material would flow into caves. Indeed, it cannot even be asserted that this phenomenon occurred frequently in rockshelters. Accordingly, the occurrence of such an event can be tied neither to a particular region nor to particular conditions [climatic?].]

Logic told Vértes that mass movement had to be the cause of these deposits. Since solifluction associated with freeze-thaw was the only process of this type of which he knew, this was how he interpreted the anomalous sediments. Note that he was not able to satisfy his own criteria for the recognition of mass movement, as no mention is made of specific structures as proof of movement. With respect to the highly rounded limestone material, see p.767 of the present text.

Since Vértes, cave sedimentologists seem to have become more and more disinclined to question the role of ice. Processes involving ice have been established as mechanisms apparently capable of transporting material into or from any part of a cave. Gelifluction deposits are claimed on the most scanty evidence.

Miskovsky states:

> La solifluxion. Celle-ci, liée à un froid intense, implique un sol gelé assez profondément pendant la plus grande partie de l'année. Pendant la période estivale seule la partie supérieure dégèle (mollisol). Ce mollisol gorgé d'eau et à l'état pâteux, coulera sur la couche gelée permanente (pergélisol ou tjäle). De telles coulées en nappes (sheet-flows [sic]) peuvent se faire sur des pentes très faibles (A.Cailleux, 1954 [reference not present in Miskovsky's bibliography]). Dans de telles couches, silex et galets seront trouvés dans toutes les positions. (1974:35)
>
> [Trans.: Solifluction, associated with intense cold, implies ground frozen to quite a depth during the greater part of the year. During the summer period, only the upper part thaws (the mollisol). This mollisol, saturated with water and in a pasty state, will flow on [over] the permanently frozen layer (the pergelisol or tjäle). Such sheet-flows

can occur on very gentle slopes (Cailleux 1954). In such layers, flints and pebbles will be found lying chaotically.]

Miskovsky is clearly unaware of the range of environments and possible deposit geometry, even given the prerequisite of permafrost. Note how he switches from a general definition to cave situations without any justification for the application of such a concept to cave sedimentology. In the rest of his text, there is not a single detailed description of a deposit that he has classed as being due to solifluction. No other more fluid mass movement phenomena, fast or slow, are recognised by Miskovsky.

Laville covers this subject in a little more detail.

> Nous rappellerons ici que le processus de solifluxion consiste en un lent déplacement, sur des pentes de faible inclinaison, de matériaux saturés et comportant des éléments fins. (1975:24)
>
> [Trans.: We will recall here that the process of solifluction comprises the slow movement, on gentle slopes, of saturated material which includes a fine component.]

Despite the fact that one might wonder why the process is restricted to gentle slopes ('What would happen if the slopes were steeper?'), Laville's introductory statement looks promising. However, he continues:

> (. . .) Il suppose un dégel superficiel du sol, une certaine élévation de la température saisonnière ou épisodique.
> Considérée généralement comme une manifestation de climat très froid, la solifluxion nous est apparue comme avoir été particulièrement active lors des épisodes que nous conviendrons d'appeler "épisodes d'instabilité climatique".
> Nous n'avons jamais décelé ses effets dans des dépôts édifiés en période de froid intense; ils semblent être intervenus à l'amorce ou au déclin des phases glaciaires. C'est là une interprétation a posteriori que l'ensemble des caractéristiques des dépôts que nous avons étudiés nous a conduit à retenir. (ibid., p.24)
>
> [Trans.: (. . .) It [the process] implies a superficial thaw of the ground, a certain seasonal or episodic rise in the temperature.
> Generally considered as a manifestation of very cold climate, solifluction appears to us as having been particularly active during periods which are most suitably called 'episodes of climatic instability'.
> We have never detected the effects [of solifluction] in deposits built up during periods of intense cold; they [the effects] seem to occur towards the beginning and end of glacial phases. This is an a posteriori

interpretation, derived from the overall characteristics of the deposits which we have studied.]

One wonders how much of a circular argument is involved here, especially since Laville, in the body of his text, in fact claims solifluction at many sites well within what he calls 'glacial phases'.

> Nous verrons que la solifuxion s'est souvent manifestée dans le remplissage des abris, par la sédimentation de matériaux d'origine allochtone.
> La mise en place de ces dépôts s'est accompagnée de caractéristiques granulométriques et morphologiques particulières du matériel sédimentaire. (ibid., p.24)
>
> [Trans.: We will see that solifluction is often manifested in shelter fills by the sedimentation of allochthonous material.
> The emplacement of these deposits is accompanied by [has resulted in?] a characteristic particle size distribution and [particle] morphology in the sediments.]

This concludes Laville's discussion in his section concerning the emplacement of sediments. However, he continues in his section on the modification of sediments:

> La solifluxion et la cryoturbation. Ainsi que nous l'avons souligné, la solifluxion a le plus souvent conduit, dans des abris sous roche, à la sédimentation de sédiments d'origine allochtone. (ibid., p.25)
>
> [Trans.: Solifluction and cryoturbation. As we have stressed above, solifluction in rockshelters has most often led to sedimentation of allochthonous sediments.]

Reading the detailed descriptions of shelter sequences, the present author found that only in a very few cases does Laville in fact offer any proof of allochthonous sediments.

> Elle s'est aussi manifestée par le déplacement de depots accumulés antérieurement; dans ce cas, elle s'est souvent assortie de mouvements de cryoturbation et il nous a été la plupart du temps impossible de différencier les deux phénomènes. (ibid., p.25)
>
> [Trans.: It [solifluction] is also manifested in the displacement of pre-existing deposits; in this case, it is often matched [?] by movements of cryoturbation and, for most of the time, we have found it impossible to differentiate the two phenomena.]

There are some difficulties here. The verb 's'assortir' currently means 'to match'; it can also have the sense 'to accompany' but all modern dictionaries qualify this sense as

as "vieilli" (obsolete). The use of the verb 'différencier' with a direct object is also puzzling. In an English translation of the introductory sections of Laville's thesis, this sentence is given as:

> It is also manifested by the resorting of previous deposits, in which case it is often accompanied by cryoturbation, and it is usually impossible to differentiate between the two phenomena. (Laville 1976:142)

As the whole of this fourteen-page translation is rather poor (cf. the rendering of "déplacements" as "resorting") and often changes the sense of the original, it cannot be taken as an accurate guide. The French original creates the distinct feeling that a certain amount of ambiguity was quite deliberately included. Does Laville mean that, where underlying deposits have been moved, it is difficult to tell whether this movement was caused by solifluction or by cryoturbation, even when solifluction was still clearly the cause of input of other sediments? Or does he mean that, where underlying deposits have been moved, it is difficult to tell whether the whole range of phenomena under discussion were caused by solifluction or by cryoturbation? In other words, is it possible that in some cases the effects of solifluction and cryoturbation may be totally indistinguishable?

In all fairness, it should be noted that here Laville is perhaps suffering from the universal communication breakdown between physical geographers and quaternary sedimentologists. The work of Tricart, although uncited by Laville, was probably the source for this discussion. Tricart writes, in an adequate English translation:

> Given debris of a suitable grain size, cryoturbation is most likely in moist conditions during the warm season. It promotes the movement of debris en masse: stone stripes, mud flows and block streams, and stratified screes. (1970:116)

Just what Tricart means is anybody's guess. This is

yet another case of ambiguity: what does "promote" mean? If Tricart is suggesting that cryoturbation causes all these phenomena, this is a far cry from the original definition of the concept (Edelman et al. 1936; see p.272 of the present text).

Returning to Laville's discussion:

> Quoi qu'il en soit, solifluxion et cryoturbation se sont manifestées par le remaniement, le déplacement ou même dans certains cas, par l'évacuation totale des dépôts. (1975:25)
>
> [Trans.: Be that as it may, solifluction and cryoturbation are evidenced by reworking, displacement or even, in certain cases, by the total evacuation of the deposits.]

Here Laville is really taxing our credulity. Under what circumstances could solifluction, as defined by Laville, produce the total evacuation of a shelter? In his section on "ravinements" (this would appear to mean any geometrical form of large scale erosion, not just channelling), Laville states:

> Le ravinement d'un dépôt peut avoir trois origines.
> - Erosion par le jeu de phénomènes de solifluxion. C'est un aspect que nous avons déjà évoqué et auquel est attribuable l'absence des horizons superficiels de la plupart des paléosols que nous avons rencontrés. (. . .) (ibid., pp.25-6)
>
> [Trans.: Channelling [erosion] of a deposit may have one of three causes.
> - Erosion by means of solifluction phenomena. This is an aspect which we have already noted and which is responsible for the absence of the upper horizons of the majority of palaeosols which we have encountered. (. . .)]

In the great majority of cases where Laville claims this palaeosol truncation by solifluction, there are no traces of the solifluction deposit itself, there is no disturbance of the underlying sediments and no sign of structures induced by ice segregation - there is just an abrupt boundary. Laville finds this idea useful to justify his otherwise untenable concept of a palaeosol (see p.530).

What then does Laville call a 'solifluction deposit'?

> Outre un agencement particulier des matériaux, décelable par la seule

observation stratigraphique, les deux phénomènes [solifluction and cryoturbation] sont identifiables en laboratoire par une distribution particulière du sediment fin, ainsi que par un modelé, un arrondi, un poli caractéristiques des éboulis.
Dans le cas de remaniements de niveaux archéologiques, les deux phénomenès ont conduit au concassage du matériel lithique: les silex taillés, totalement défigurés, présentent le plus souvent une forme émoussée et des pseudo-retouches alternes épaisses (Pl.9, n°3).
(ibid., p.25)

[Trans.: In addition to a characteristic arrangement [structure] in the sediments detectable by stratigraphic observation alone, the two phenomena [solifluction and cryoturbation] may be identified in the laboratory by a characteristic size distribution in the fine material, as well as by a characteristic relief, rounding or polish [wear] on the scree particles.
When archaeological levels have been reworked, the two phenomena have produced crushing of the lithic material: flint artefacts are badly damaged and usually have a blunted [worn] form and deep, alternating pseudo-retouch (Plate 9, no.3).]

Laville does not define the particular fabrics and structures visible at his sites at this point in the text, nor does he describe them in the detailed sections for individual sites, except in the case of a few clear cryoturbation (sensu stricto, see p.272) structures. He does not seem to be aware of any fabrics which indicate lateral movement, as is further suggested by the ambiguity concerning the difference between solifluction and cryoturbation. The only criteria that we can be sure he is using are poor sorting, (apparently) mechanical abrasion of limestone and damage to flint artefacts, and rarely all three of these in any individual case as reference to the body of his text will show. Even the photograph (ibid., Plate 9, no.3, p.34) of artefacts damaged by cryoturbation shows pieces, not from a site where they are demonstrably in association with observed cryoturbation structures, but from the old archaeological tip at the upper shelter of le Moustier (Dordogne).

As a generalisation concerning the literature on cave and shelter sites, it may be said that not only is any sort of proof of the involvement of ice in any form nearly always lacking, thus

precluding any strong claim for gelifluction, but the dearth of observations of orientation, fabric and deposit geometry often leaves the reader with the impression that there is no real proof of any kind of mass movement at all. In most cases, the lower boundaries of these deposits are abrupt, thus we must be dealing with overriding masses. The deposits have sometimes been followed downslope for distances of ten metres or more, distances which could only be covered in about three millennia. Judging from the statements by these authors concerning chronology, such time scales are usually much too long; the build-up of autochthonous sediments would have swamped such masses almost as soon as they transferred anywhere near a cliff face. Even allowing that these are indeed solifluction deposits, the claims that they are responsible for massive erosion, including truncation of palaeosols, cannot be accepted. In this author's opinion, the vast majority of reported solifluction deposits in caves and shelters are to be attributed either to phenomena not involving lateral mass flow or to phenomena associated with fast mass movement. Despite the fact that he too has misused the term 'solifluction' in the past, the author has not yet seen a really convincing solifluction deposit, let alone a gelifluction deposit, in a cave or shelter whether in Britain, France, Belgium or Spain. If solifluction is indeed a rare event, the environmental and taphonomic inferences usually drawn from the occurrence of badly sorted cave sediments may be suspect in the extreme.

9.3.4. Mass Movement Involving Very Fine Grained Sediments

Sediments with very high proportions of colloids and fine silts may behave rather differently from other material.

In coarser sediments, much of the strength is due to interparticle friction; they are said to be cohesionless. As the fine component increases, cohesion becomes more significant. Cohesion is the result of a number of forces acting to produce a net attraction between fine particles. These forces include: interparticle gravitational attraction (weak); van der Waals bonding (alternations in the number of electrons at any one time on one side of an atomic nucleus - weak); ionic bonding ('lending/borrowing' of an electron to achieve a stable outer electron shell - medium); covalent bonding ('sharing' of an electron pair - strong); hydrogen bonding (a two-part bond, involving a hydrogen atom covalently bonded to one atom and ionically bonded to another - medium). Note that the term 'cohesion' is also applied to the 'bonding' of smaller particles by water surface tension; this is only important if the sediment is damp but nowhere near saturated. In most sediments, which are more or less graded, the need to overcome both interparticle friction and cohesion results in the clear distinction between slow and fast mass movement. In most fine sediments, friction will be less important. Failure may still occur, in various forms, but it is localised, occurring along well defined planes. Cohesion within the sediment mass continues to be of great importance even after failure. The result is that mass movement in such sediments may occur at 'slow' or 'fast' speeds, but also in the 'medium' velocity range not usually seen with other sediments. The actual speed will be a function of (amongst other things, such as slope) the degree to which cohesion has been reduced. Most real cases of interest to us will be nearer the slower end of this range, a fact which allows the inclusion of this theoretically difficult group in the present section. The forces of attraction noted above can be

reduced, before any movement occurs, mainly by factors associated with weathering; indeed, the breaking of covalent bonds cannot be achieved without chemical alteration. However, perhaps the most important feature is how the sediment cohesion reacts to initial movement, often called 'remoulding'. This characteristic is termed 'sensitivity'; highly sensitive sediments quickly lose their cohesion once they start moving, so that the movement rapidly escalates.

The terminology used to describe the movement of fine sediments is rather confused. The most common type will be called mudslide here. The same phenomenon is also referred to as '(slow) mudflow' by many authors (cf. Statham 1977), but it is probably best to avoid this term as it has been so misused in the past. All authorities are agreed that both sliding and flow are involved, and that sliding plays a much more important part than in most other forms of mass movement. The term 'mud' implies here a dominantly fine sediment and not just any sediment with a wet, 'muddy' matrix. Larger particles may of course be included in mudslides, but the particle distribution is always markedly bimodal, with a relative gap between the major mode (fines) and the minor mode (coarse elements).

Mudslides usually originate as rotational slips on high angle slopes. Slipping tends to maintain the high angle, so that, even if the original slope was provided by a small feature such as a low bank, there may be considerable headward erosion across areas of little or no slope. Large scale erosion forms in cliffs of fine sediment have been called 'corries' (Hutchinson 1970). Depending upon its cohesion characteristics, newly slipped material may require further softening by mechanical or chemical alteration before it will flow and slide further. Softening

proceeds until a significant proportion of the material has reached a plastic state. However, there may still be many hard remnants of the original deposit (lithorelics) that will never become soft and which will survive relatively undisturbed to be recognised in fossil deposits. Such lithorelics may be of any size, but small units might appear very like the aggregate-cored 'clay clumps' recognised by Goldberg in supposed soliflucted material (see p.349). The softened mass will flow out from the source, producing the classic viscous flow morphology, including 'levées' and lobes. Flow itself may be rather slow, reflecting the remaining cohesion of the fine sediment. Nevertheless, overall movement is observed to be jerky, showing the influence of sliding; it is this feature that gives average velocities of c.10-20m/yr, with short pulses of 100m/yr or more. Major shear planes are usually to be found near the base of the mass, but there may be other planes at higher levels as well, reflecting the stacking of individual lobes. High angle shearing, both with distal material slipping away from proximal material (normal shearing or faulting) and with proximal material riding up over distal material (reverse or thrust shearing or faulting), may also occur. Shear planes at any angle may survive as 'slickensides' (very smooth and polished surfaces) in fossil deposits.

Since mudslides are so often included under different headings, such as solifluction, there is little direct information concerning many of the topics of interest here. Fabric orientation will indicate the flow pattern. Erosive capability would seem to be concentrated at the headwall, since there is no obvious mechanism for basal erosion. Carrying capacity for large objects should be appreciable, since the 'mud' is both dense and cohesive. Because deformation (flow) is minor, fragile objects

should suffer little damage. Because of the slow rate of water loss to be expected with fine materials, mudslides might perhaps mould themselves more exactly to particulary tortuous terrain (eg. a winding and branching cave passage).

For a mudslide to occur a special type of sediment is required but, given this prerequisite, mudslides may occur in any environment where water is available. In temperate environments, rainstorms usually provide the water, whilst in cold environments, thawing of ice and snow is the main source. This explains the observations of Chandler (1972; see p.333 above) and others concerning fine grained deposits, which are environmentally gelifluction deposits but mechanically mudslides. Naturally, such features as permafrost and ice lenses will encourage mudslides to occur, given the right sediment. Indeed, mudslides on slopes below about 4° have rarely been observed except in areas where ground ice can encourage sliding. In cold environments, the possibility of lithorelics may also be increased by the fall of frozen blocks of sediment from the retreating headwall. Such blocks may sometimes survive relatively undisturbed in fossil deposits and, if they contain sediment that would not have enough cohesion to hold together on its own, they may provide useful indicators of frozen ground. It should be noted, however, that thaw slumping (Washburn 1979), creating large shallow erosion scars, although common in fine sediments is not totally restricted to such sediments. With coarser or better graded deposits, the slumped debris may be removed by true solifluction, slopewash or even fast mass movement (cf. French 1976).

Typical mudslides have not yet been explicity described from caves despite the fact that there is no reason to suppose that they might be unlikely in such situations.

Quick clay flows represent the fastest type of movement in fine sediments; velocities of over 40km/hr have been recorded. Such flows are totally source-dependent and, as such, are rather rare. They are described briefly here, not because they are very likely to occur in caves, but because the causes of the sensitivity of these sediments graphically illustrate the concept of cohesion. Two causes, which appear equally probable either separately or in combination, have been proposed. Rosenqvist (1953), working with emerged marine deposits, suggested that the 'clays' derived their structure and cohesion from the fact that they had been deposited under saline (electrolytic) conditions. As subaerial leaching progresses, salinity drops and there is no longer any substantial bonding to support the delicate flocculated structure. Flocculation due to an electrolyte like sodium chloride produces a structure in which colloids and silt particles are intimately associated, a process known as symmixis (cf. Pettijohn 1975). Such a structure is of course totally unstable once the electrolyte is removed. Given conditions for an initial rotational slip, the slipped material is remoulded and immediately loses practically all of its cohesion. Cabrera and Smalley (1973) suggested that, as many of these sediments have an ultimately glacial source, much colloidal quartz and feldspar would be included with a reduction in the proportion of true clay minerals. Compared with the latter, quartz and feldspar are more or less electrochemically inactive and do not usually participate to any great extent in any of the medium or weak bonding situations noted above. Sediments with a high glacial input will therefore lack cohesion once remoulded. This suggestion is of great interest, since many mass movement phenomena, slow and fast, involve glacial sediments. Any percentage of electrochemically inactive colloids

should have a proportional effect upon the mobility of such sediments. Note that glacial sediments *in situ*, as well as sediments deposited under significant heads of water, are intrinsically unstable as they are overconsolidated (i.e. the pressure under which they were formed has since been removed, leaving them compressed rather like a coiled spring).

9.4. Fast Mass Movement

9.4.1. General

This section will be divided into two parts. The first deals with phenomena in which an element of flow is obvious. The second deals briefly with extremely high speed phenomena in which free-fall, flow and/or air-cushion gliding may be involved.

9.4.2. Debris Flow

The deposits resulting from debris flow have rarely been examined in detail in the cave situation, despite their common occurrence. The subject has, however, been studied in open sites, especially in connection with the formation of alluvial fans. Statham gives a general description of this process:

> One style of flowslide which has been noted in widely different climatic regimes and sediment types is the *debris flow*. Debris flows consist of a fluidized lobe of poorly-sorted coarse and fine sediments which originate on debris-mantled hillslopes. The lobe moves at a few kilometres an hour, leaving low ridges of debris or levées . . . on either side of its track as it goes. (. . .) The track itself may or may not be eroded by the passage of a debris flow; . . . . (1977:93)

Theoretical considerations of the various forces involved have been undertaken by many investigators (cf. Johnson 1970; Fisher 1971; Rodine & Johnson 1976; Pierson 1981).

Examination of the literature shows that debris flow is a complex process, with several different styles depending upon local geomorphology. In order to obtain information as closely applicable to the cave situation as possible, we will concentrate on flows that, during some part of their descent, are restricted to channels.

Three zones of activity may be defined: (a) higher, steep slopes covered with debris, from which much of the sediment is derived by slippage or very rapid flow when the finite yield strength of the sediment is exceeded, usually after massive and sudden intake of water; (b) a middle zone of moderate slopes where flow is often confined to rock- or sediment-cut channels; and (c) a zone of low gradients where flow may be concentrated in ephemeral channels but, more often, is relatively unconcentrated and present as lobes and sheets (eg. the outer margins of an alluvial fan).

The angle of slope required for initiation of debris flow depends largely upon the sediment type; reported angles vary from near vertical to a minimum of about 20° (cf. Statham 1977). The slope angle at which debris will continue to flow may be as little as 2° (Johnson 1970; Pierson 1980). Initiation of flow is always linked with the sudden availability of large quantities of water. The trigger mechanism may be high, but not necessarily exceptional, local rainfall (cf. Rapp 1960; Prior et al. 1970) or rapid melt, or partial melt ('slush avalanches'), of snow and ice (cf. Rapp 1960; Jahn 1967; Johnson 1970). Climate is only important in as much as it may be considered that the trigger mechanisms are more or less likely to occur, or occur frequently, under given climatic conditions. Soil formation including a vegetation mat should reduce the frequency of debris flow events

but might also ensure that they are more catastrophic when they do occur. Debris flow has often been observed in areas with a frozen substratum. In this case, the distinction between debris flow and gelifluction has, in the past, been to some extent arbitrary (supra). Debris flow is here considered to be faster, usually by several orders of magnitude, with initiation involving failure; continuation of the flow is not necessarily dependent upon the lateral persistence of the pergelisol.

At least in zones (b) and (c), as defined above, observers report two distinct phases in the development of debris flows: (1) the true debris flow phase, when comparatively viscous material with a high solid content is flowing downslope, often as discrete pulses; (2) a liquid phase, during which very muddy water is flowing (cf. Pierson 1980). Temporary dams of debris often block channels, ponding water behind them; when the dams eventually rupture, the muddy water runs on ahead of the more slowly moving debris. Debris flow ceases when the combined effect of decreasing gradient and decreasing water content causes the flow to 'freeze' as internal frictional strength overcomes the kinetic forces. For a given sediment and a given gradient, this limiting point may be expressed as a critical thickness of deposit, below which it will no longer flow.

The amount of sediment transported in debris flows is obviously strictly a function of local conditions. Dry environments with heavy seasonal rainfall seem to produce most of the larger flows: R.H.Jahns (cited in Johnson 1970) observed a flow in motion with a snout 35 feet high, whilst J.S.Shelton (cited in Johnson 1970) recorded an inactive lobe, 2000 feet wide and 50 feet high. The upper velocity limit is similarly a function of slope, water content, sediment type, etc. Pierson (1980) records a peak velocity

in channelled debris flow of 18km/hr, corresponding to a debris discharge of 20m³/sec. Much higher but short lived velocities may develop near the source of the system if slippage as well as flow is involved. In some materials, with a very fine matrix, slower flows may occur, eventually grading into mudslides.

Observations concerning the deposits and deposit geometry which result from debris flow are extremely varied and causative factors are sometimes obscure. All observers agree that the deposits are nearly always poorly or very poorly sorted; Pierson (1980) points out that this is necessary to prevent particle interlocking. Coarse elements, if present, float in the fine matrix. Reineck and Singh (1980) report poor graded bedding in more fluid flows but lack of such structure in viscous flows, whilst Fisher (1971) reports inverse graded bedding. From the plates in Fisher's article, this feature would seem to be a zonation between coarser and finer particles rather than continuous grading. It seems probable that many of the reports of graded bedding reflect a more general phenomenon which leads to sorting out of coarser particles towards all the boundaries of the flow; usually, authors who record graded bedding are either definitely referring to unchannelled flow or are making general statements with no reference to sediment geometry (but see p.379). Whenever detailed information is available, it can be seen that coarse-tail grading, rather than distribution grading (Middleton 1967), is the rule. The coarsest material is progressively lost upwards (normal), or downwards (inverse), and the particle size distribution does not shift but is merely truncated. This is also a feature of subaquaeous, high solid concentration turbidity currents.

Sharp (1942) named lateral ridges of coarser material "levées" and considered that they formed due to a 'snow-plough'

effect at the snout; such a process has been confirmed by Statham (1976a). However, Johnson (1970) and Pierson (1980, 1981) are convinced that a sorting process is also active within the flows (infra). A statement by Brunsden would seem to be somewhat misleading: "The lobe comes to rest when it drains both at its margins, providing sinuous levées of coarse material, and . . . ." (1979:175). That water rapidly draining from a deposit can remove fine material, and thus leave behind 'cleaner' coarse debris, is certain. However, levées are in no way formed by draining water, as a simple calculation of required water velocities, and the extreme deceleration of the water immediately after it left the mass of debris which would be necessary, will show. Johnson (1970) reports coarser debris, carried near the middle upper surface of flows, which is moving faster than the flow mass; this debris therefore tips over the snout and is reabsorbed at the base, in the manner of 'caterpillar tracks'.

Fisher (1971) records strong pebble orientation parallel to the base of the flow; Reineck and Singh (1980) confirm this in more fluid flows but report random and even vertical orientation in more viscous flows. Blatt, Middleton and Murray state:

> Particles of non-spherical shape freely suspended in viscous shear flows tend to rotate continuously. The rate of rotation varies according to the orientation with respect to mean flow so that particles have a relatively high 'residence time' in orientations close to the flow direction. (1980:122)

Although it is not clear under what exact conditions the above observation was made, the present author has noted this behaviour in experimental debris flows. As the density (number) of coarser particles increases, the tendency towards 'horizontal' bedding is accentuated due to interference between particles. The fine matrix of debris flows is often massive. However, pseudo-vesicular structure, caused by entrapment of air, has

sometimes been noted (cf. Reineck & Singh 1980).

Blatt, Middleton and Murray (1980) are the only authors amongst those cited above who have recognised the presence of imbricate structures in the coarse debris. 'Imbricate structures' refers here to clear lines of pebbles, nose resting on tail, not merely to a generalised average dip of coarse particles. These authors use this observation to question the generally accepted opinion that such structures indicate moderate to strong current (i.e. non-viscous flow) conditions in marine or fluvial settings (cf. Pettijohn 1975). These two points of view are not necessarily irreconcilable. The present author suggests that imbricate structures, developed in obvious debris flow sequences, might be due to the 'muddy water' phase noted by many observers. At Pontnewydd (see p.1192), such structures occur only at the boundaries of separate flows and never embedded within a single sediment body.

Despite many general statements that debris flows may be entirely non-erosive, it is clear from the primary literature that, when confined within channels, debris flows may be moderately or even radically erosive (cf. Statham 1976a). Pierson (1980) reports up to ten metres of vertical entrenchment in only a few days; he notes that it is the true debris flow phase, not the associated 'muddy water' phase, that is primarily responsible and sites corrasion (scouring) as the process involved. Johnson (1970) considers the channels eroded by debris flows to be typically U-shaped in cross section. The 'snow plough' effect noted above must also be somewhat erosive.

Given the diversity of observations on debris flows, it will be instructive to consider the main points of theoretical rheology that have been suggested to account for them. Sharpe

(1938) considered these flows (included under the heading of "mudflows") to be due to continuous plastic deformation. Johnson (1970) has proposed a more complex Bingham viscosity model, combining aspects of Newtonian viscosity and plastic deformation, which seems to fit his experimental data quite well. Note that both the above models predict that velocity would be zero at the flow boundaries, a condition which cannot hold in reality. Such a flow could not be erosive by definition, and observations recording the passage of debris flows with neither erosion nor deposition (cf. Statham 1977) must reflect basal shear. Nevertheless, such models may be useful away from the boundary conditions. One of the results of the Bingham model is that velocity would increase away from channel sides until a stable velocity is achieved; a central "plug" of material, moving at maximum velocity but with little or no internal relative velocity, should result. Johnson uses this concept to explain how fragile objects (eg. very heavily fissured boulders) can be transported within debris flows without damage. This might also help to account for the maintenance of discrete pockets of different sediments, although laminar flow at low shear rate gradients should accomplish the same end. Johnson's model requires that flow be laminar; Fisher (1971) supports this conclusion based upon his observations of parallel bedding of coarse elements. However, it is difficult to envisage truly laminar flow near irregular channel walls or where corrasion is actively eroding underlying sediments. Even gently flowing water does not fulfil these theoretical conditions exactly. Pierson (1980) has observed turbulent flow, indicated by downstream surface V-waves, at the height of each debris flow pulse; he states that lateral sorting of coarse debris is most active at the transition from laminar

to turbulent conditions. Just why V-waves should develop is not clear; turbulent flow should produce a velocity profile rather like the Bingham model. If turbulence is so important as to set up major riffles (longitudinal spiralling) and other vortices, there is little wonder that coarse material is sorted to the margins. If the concept of Reynold's Number is applicable to such heterogeneous flows, with a given velocity and kinetic viscosity, turbulent flow will develop in a U-shaped channel more easily than in a channel of any other shape with the same cross-sectional area. Pierson (1981) also accepts Johnson's (1970) proposal that sorting may be achieved during laminar flow by dispersive pressure, involving migration of larger particles towards zones of lower shear rate. Johnson has suggested that such a process should also produce a central core of coarse debris, which he likens to medial moraines in glaciers; it is difficult to accept Johnson's mathematical analysis on this point. Blatt, Middleton and Murray (1980) have pointed out that chaotic fabric should be characteristic of any central 'plug', but this is only acceptable if the 'plug' suffered practically no deformation at any stage during the movement.

One other form of debris flow may be mentioned briefly. The geologic record contains comparatively rare occurrences of 'tilloids' that have been deposited as subaquaeous flows (cf. Pettijohn 1975). Individual beds may be tens of metres in thickness. Sliding and plastic deformation are usually involved and may even be dominant over true viscous flow. Erosion may be extreme, especially where slope suddenly increases. Because of constant total saturation reducing internal friction to a minimum, such flows are unstable and will easily be transformed into very high velocity turbidity currents. Most tilloids of this type

contain some zones of good graded bedding (usually, but not always, coarse-tailed), thought to be indicative of minor current events involving lift by turbulence. Very similar phenomena have been reported, on a much smaller scale, from lacustrine environments (cf. Reineck & Singh 1980).

Despite the availability of information from various environments, Brunsden was still able to write in 1979:

> Yet they [debris flows] are among the least studied and understood of all mass-movement phenomena. There is no satisfactory classification, there is much confusion in terminology and the mechanisms involved are poorly understood. (loc. cit., p.173)

However, a new paper by D.E. Lawson (1982) may possibly go a considerable way towards setting up a general model. Indeed, there is much here that has a bearing on all forms of more fluid mass movement. For this reason, Lawson's marvellously detailed observations have not been fragmented for inclusion in the above discussion, but are set out below as a whole.

The observations involve material which, until Lawson's work, was usually included under the heading of 'flowtills'; the deposits occur near the end of the Matanuska Glacier in south central Alaska. This surprising study area is in fact ideal, because the abundance of summer melt water and sediments ensures frequent debris flow events with the great majority of combinations of contributing factors. The source sediments are either morainic (sensu stricto) or are associated with stagnant ice (melt-out, mass flow, fluvial and lacustrine sediments); the particle size distributions are thus extremely varied, although silt is usually a significant component.

> Mobilisation processes generally mix these materials into a very poorly sorted, gravelly sandy silt. The ability of a flow to transport larger particles determines if any selective sorting occurs at flow mobilisation. At some locations for example, unsupported boulder- and cobble-sized clasts remain in the source area as a lag stranded on a bed of ice. (D.E. Lawson 1982:281)

Three modes of flow initiation are recognised: (1) rotational slumping of sediments overlying stagnant ice, causing remoulding and flow; (2) backwasting (cf. thaw slumping, p.341 above) of sediment and stagnant ice, with shearing, mixing, remoulding and absorbtion of meltwater, eventually resulting in flow; and (3) very rapid ablation of 'dirty' ice, with immediate flow.

Lawson's sections on movement and deposition contain so much significant information, especially if one has a synthetic interest, that they might almost be cited verbatim. However, in order to reduce the sheer bulk, this information will be given in note form here, although this does Lawson little justice. Note that statements concerning internal behaviour during movement are based on actual observations carried out from within a plexiglass 'cage'.

Type I Flows

Low water content (11-17% dry wt.). High bulk density (2.0-2.6g/cm^3). Lobate flows; blunt head higher than main flow surface, but continuous with marginal ridges; marginal and frontal slopes >40°. Body with planar surface and constant thickness, thinning abruptly upslope in the tail. Compressional deformation (resistance at flow margins) gives arcuate ridges paralleling the head configuration. Also ridges due to exceptional input of sediment at headwall (source). Max. dimensions: 2m thick, area 15x45m. Surface flow rate: 0.1-0.5cm/sec on slopes of 1-10°, but rarely 0.1m/day or less. Max. travel distance observed: 300m/28days. Texture generally homogeneous; matrix-dominated (pebbly sandy silt) with suspended cobbles and boulders. Bulk mean grain size: 2.0-0.3mm. Heads may be coarser. Water content, density, porosity and shear strength generally constant in main part of flow. Shear strength at centre surface: 0.4-1.5kg/cm^2. Air-filled voids (partial water saturation only). Unconsolidated lithorelics with random orientation. No preferential orientation of pebbles; mixing during mobilisation generally destroys fabric of source material. Lack of internal deformation during movement; shear concentrated in first few centimetres only at base. Head of flow continuously overrides basal shear zone; sometimes upper material rolls over and down, producing tension fissures. Non-erosive (therefore not channel-creating) but loose materials in path may be incorporated or pushed ahead. Underlying incompetent sediments may be deformed (décollement folds) or may fail along high angle thrust planes; incorporation may follow. Overriding may cause rapid water expulsion or extrusion of underlying cohesive (plastic) sediments. Morphology and dimensions maintained during final deposition. Strength of surface materials resists reworking by meltwater; even true stream

flow results only in a coarsening of the surface.

Type II Flows

Water content: 17-24% dry wt. Basal shear zone thicker than in Type I, giving a central plug of unsheared material; often channellised, with lateral margins as well as base in shear. Flow rate: 0.2-5.0cm/sec. Lobate head only slightly higher than main flow surface; quite low angle snout. Plug surface usually smooth, covered by silt-laden meltwater. Water may erode braided rills with fine gravel lags. Possible arcuate ridges, sometimes with lithorelics. Snout material may be coarsened by meltwater entrainment. Thickness and other dimensions decline markedly with increasing water content. Max. thickness: 1.6m; average: 0.2-0.7m; min.: 0.01m. Channel width depends upon surrounding material; ice gives narrow channels (min. 0.5m), sediments near failure give wide channels (max. 5.0m). Bulk mean grain size: 1.9-0.1mm; but texture different in different parts of flow. Material of plug reflects source, texturally heterogeneous and poorly sorted. Lithorelics may be added to plug from channel walls during movement. Sheet flow of water at surface may give thin silt deposits. Levées possible. Lower shear zone has a layer of tractional gravel, occasionally with coarse-tail grading. Upper shear zone intermediate between base and plug; better sorted than plug if water content higher (more sand + coarse silt). Central plug non-deforming, without preferential orientation; shear strength: 0.6kg/cm^2 to unmeasurable (inversely proportional to water content). Saltation and collision of coarser particles in lower shear zone may produce dispersive pressure sorting. Very large particles may have their upper parts trapped in, or in contact with, the plug; they are dragged or rolled along. Brief periods of "turbid" [= 'turbulent'?] mixing at bed irregularities or sudden increase of slope; even plug material and surface water may be mixed, causing loss of strength and thus 'settling' of coarsest particles. If marked difference in water content between plug and shear zone (the latter being under excess pore water pressure), fluid expulsion tubes (1-2mm diam.) may traverse plug, usually in clusters. Fines normally lost to water flow at surface, thus coarsening shear zone sediments; pockets of coarser particles so formed may 'settle out' to lower shear zone. Flows may occur as surges, with intervening episodes of water flow. Potential for preservation of structural features in the area of final deposition drops as water content rises. Tractional gravels may be dumped upslope of the main depositional area. Meltwater may either deposit fines over, or erode, mass flow deposits.

Type III Flows

Water content: 25-33% dry wt. Shear throughout. Flow channellised at rates of 0.15-1.25m/sec. Thickness (max. 0.6m) decreases regularly from head towards tail. Irregular surfaces often due to protruding tractional material. Texture homogeneous (continuous mixing) at any one point along the flow with coarse particles throughout, but some concentration of fines near base. Bulk mean grain size: 0.5-0.06mm; more gravel at higher slope angles. Marked size sorting along flow, with tail finer and snout coarser, the latter due to bulldozing, entrainment of coarse channel bed material, meltwater winnowing. Flow occurs as surges, with intervening water flow. Frequent temporary dams (coarse material); release causes turbulent mixing and concentration of sediment in the head region. Highly erosive. Sloshing, at bends or abrupt changes of slope, gives thin levées or overbank deposits. Coarser particles observed with long axes either parallel or at right angles to flow direction. At final deposition, drier flows may lose coarse tractional load first, then the body of the flow, lastly surface fluids and fines.

Wetter flows may give a more regular pattern, fining distally. Water flow usually coarsens deposit surfaces. One flow 'episode' usually comprises several separate surges and depositional sequences.

Type IV Flows

Flows are source dependent: fast melting debris-poor ice. Water content: 35-70% dry wt. Low bulk density: min. 1.1g/cm^3 (i.e. close to that of water alone). Bulk mean grain size: 0.1mm; homogeneous sandy silt or coarse-silty sand. Coarse sand and granules sink immediately when placed on flow surface. Surface flow rate: 0.01-2.00m/sec; rapid reaction to changes in slope. Erosive. Channels less than 1.5m wide. Thickness: 0.02-0.15m. Some small pebbles in traction. Parallel lines of movement on surface [longitudinal, said to indicate laminar flow]. Final deposition only occurs on very low slopes, thicker deposits being restricted to depressions. Fluid escape structures; massive, or normal or coarse-tail grading; no preferential coarse particle orientation. Sediments often incorporated (before deposition), or reworked (before total loss of water), by water flow.
(summary of D.E. Lawson 1982:285-97)

Lawson's observations clearly illustrate the continuum of relationships between the various flow parameters and the physical evidence that we may expect to be able to recover from fossil deposits. In general, we may say that as one set of variables increases (water content, surface flow rate, typical length of flow, porosity, erosiveness) another decreases (flow thickness, bulk wet density, bulk mean grain size, shear strength). Fabric development is also an increasing variable, although flow fabric is lost when Type IV flows come to rest. It is the recognition of these relationships which is important here. The actual empirical definition of the four Types is not generally applicable, nor is the exact numerical relationship between variables. The reasons for this fall under two headings: (1) many of these flows were underlain by ice, which provided a slip surface and additional meltwater during movement; (2) because of the glacial source, low cohesion was to be expected, due to the relative lack of true clay minerals, organic matter, etc. When dealing with fossil deposits, we do not need, or even desire, a classification scheme. All that is required is detailed observation of the deposits and a general understanding of how these observations relate to the

original flow parameters.

As was noted at the beginning of this section, very little is known about debris flow in caves. Two factors, particularly characteristic of caves, should be of importance. First, few caves show simple slope profiles. Commonly caves are made up of alternating near horizontal passages and near vertical pitches; passages may even rise locally creating sumps. In plan, caves are also generally very much more tortuous than average surface channels. Boulder chokes and collapse blocks may effectively render a cave even more irregular. Consequently, the relatively simple pattern of erosion and deposition seen in open sites will have little chance to develop within a cave. The second factor of importance concerns the extremely slow draining of debris flows to be expected in most caves. Pre-existing sediments will usually be near to saturation before the arrival of the debris flow and the cave air will also be close to 100% relative humidity. Water can therefore escape only in the liquid form, and only from free surfaces of the debris flow deposits. With long term maintenance of the fluid, or at least plastic, state, local cave topography should be the main controlling factor in the development of debris flows.

Savage (1969) has described the events at G.B. Cave (Mendip) which led to the formation of a major debris flow. The sediments have still not been examined in detail, but several observations are strikingly reminiscent of features discussed above from open sites. Note that the site lies only slightly above normally active streamways, so that flooding was also involved. The event was triggered by intense local rainfall. Lacustrine deposits in a small closed depression collapsed into the underlying cave, incorporating much coarse limestone debris on their way. Further collapse was noted during the winter after

the main event and it seems likely that this will continue for some years to come, despite the presence of the old cars which have now been used to block the dangerous shaft. When the present author first visited the site in 1981, one side of the deep conical collapse cavity showed slumping with plastic deformation, but most of the perimeter was bounded by near vertical shear surfaces on which totally undisturbed lacustrine bedding was clear. Within the cave, the resulting debris flow (or flows) occupies one main topographic unit, extending up to 100m from the point of entry. Savage's paper was published in March, 1969, some eight months after the initial collapse; events of December, 1968, are mentioned. He reports that at the time of writing large boulders were still moving in the "thick thixotropic mudflow" (ibid., p.124), indicating that the deposit had not been able to drain. This also illustrates the concept of critical thickness; the overburden of a boulder fluidises the matrix in the immediate vicinity and allows the boulder to move downslope, without the whole deposit moving as well. When the present author revisited the site in 1982 (thus 14 years after the initial event), some areas of the deposits, both with boulders near the source and without coarse elements distally, continued to be dangerously thixotropic ('quick'). Massive erosion of pre-existing deposits has taken place, at least beyond the observed debris flow. Much of this erosion was probably due to the large amounts of water which penetrated the system, both through the collapse cavity and through other surface links in the vicinity. It is tempting to visualise the possible destructive power that miniature 'turbidity currents' might develop under these circumstances. The possibility of subaquaeous flow or slide has also been suggested by Drew and Cohen (1980) for a fossil deposit in an

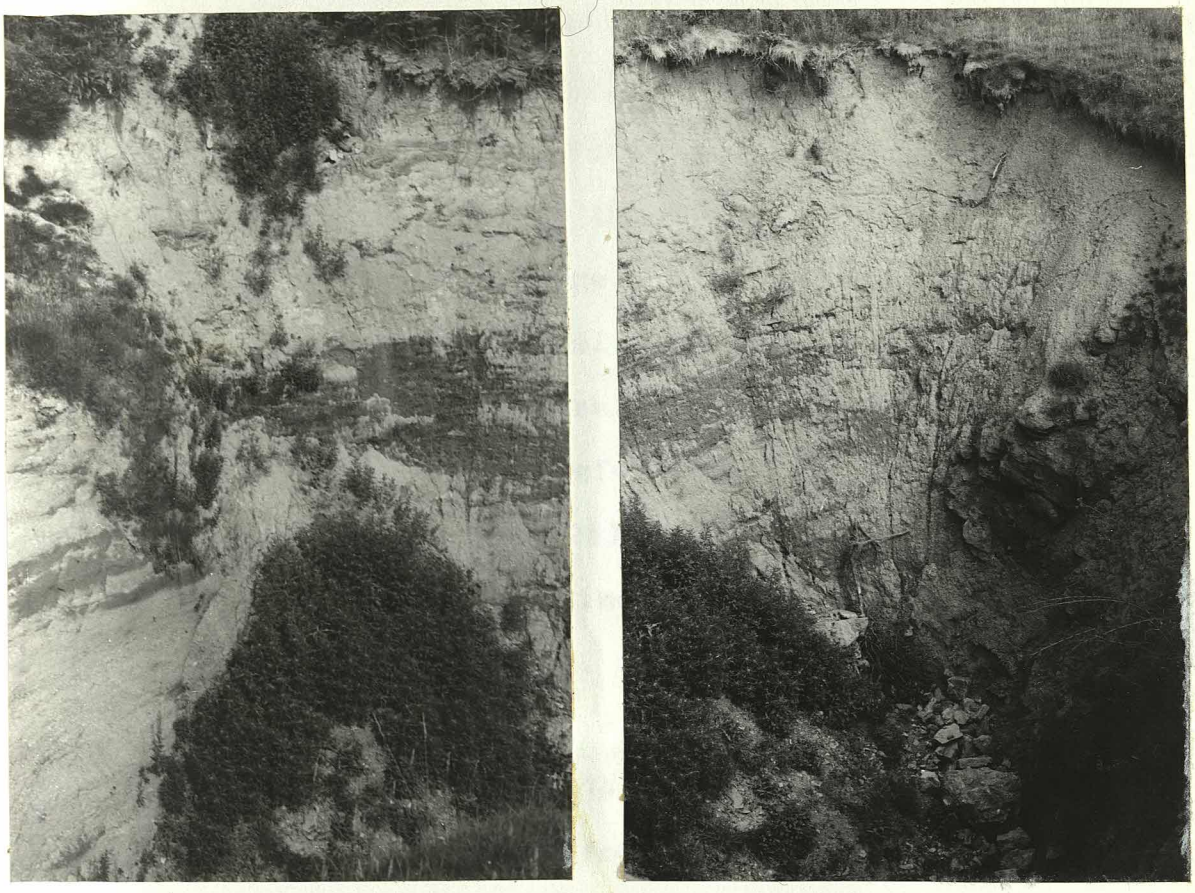




Fig.24 G.B. Cave - Exterior Collapse

Note shearing in some zones, revealing horizontal lacustrine bedding, and chaotic slumping in other zones.

Irish cave. The debris itself at G.B. Cave has now been massively eroded by water flow; in 1982, we were able to recognise debris flow material stuck in wall crevices up to 2m above the present stream bed. High remnants like this seem to be more common at constrictions in the cave passage, suggesting that debris built up higher at these points than usual. It is to be hoped that data on the fabric and geometry of the G.B. Cave debris may become available in the near future, before the stream removes all but the large rocks (see figs. 24 - 26).

Debris flows have sometimes been recognised more or less explicitly in caves in the past. Till-derived material in unsorted deposits well within a cave may often indicate debris flow (eg. Cefn Cave, Boyd Dawkins 1874; Cae Gwyn Cave, Hicks 1886a; Elderbush Cave, Bramwell 1964; Pontnewydd Cave, see p.1181). Exotic material derived from alluvial gravels may be included in debris flows, as in the lower deposits of Joint Mitnor Cave (see p.806) and Pixie's Hole (see p.842), both in Devon. It is more difficult to recognise a debris flow if the source is purely local; Banwell Bone Cave (Mendip) probably contains several such flows. A minor flow, this time in the upper levels at Pixie's Hole, reached a point within the cave (c.30m from the present entrance) where the slope changed from 5-10° to c.35°; it then channelled the underlying late pleistocene deposits to a depth of almost a metre, and with over five metres of headward erosion, redistributing fauna and later upper palaeolithic artefacts up to at least 50m into the system. At Three Holes Cave (Devon) (see p.766), a debris flow is exposed at the back of the cave that contains masses of fauna, mostly carnivores, traces of a lower palaeolithic industry and large numbers of truly spherical limestone cobbles. A complex series of debris flows at Pontnewydd

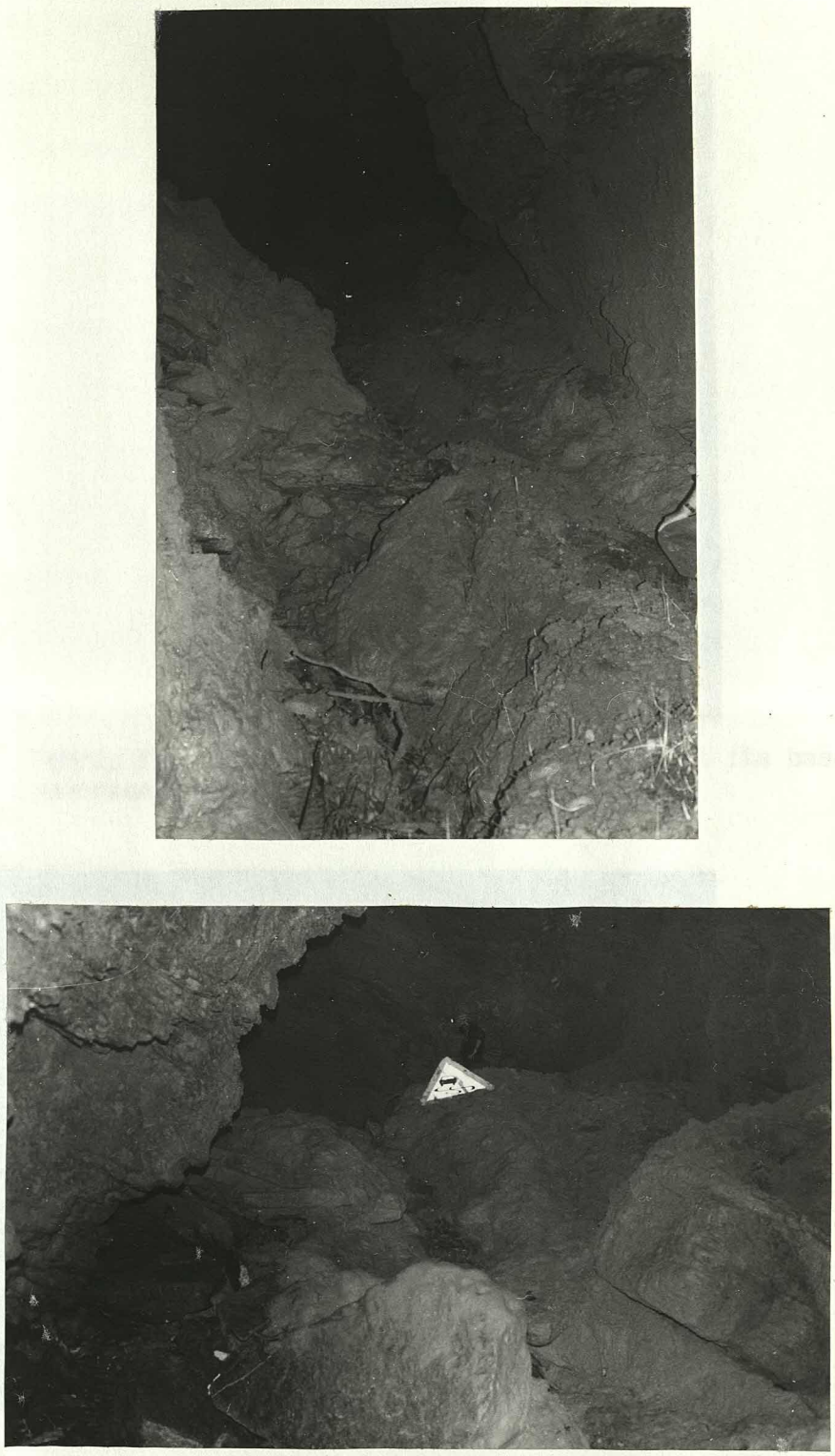
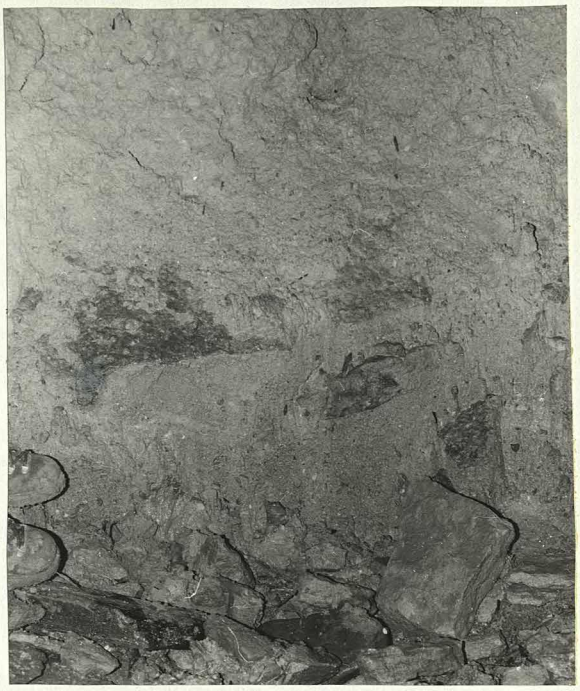


Fig.25 G.B. Cave - Interior Debris Flow

Note the huge slabs moved by the flow and the abundant introduced organic matter (including seeds giving rise to etiolated plants). The sign no doubt represents a local caver's prediction concerning the old cars which have been dumped into the entrance in the vain hope of blocking it.



Debris flow incorporating convoluted sands at its base, reworked from pre-existing stream deposits.



[horizontal]

after-flow

debris flow

Note the sorting of larger rocks and branches to the bottom of the debris flow; the well bedded current deposits of the after-flow; and the 'flame' (fluid escape) structure [arrow] indicating that the debris flow was still uncompacted and partially fluid even after considerable after-flow deposition. The section is in the side of the active stream channel.

Fig.26 G.B. Cave - Details of Debris Flow

Cave (Clwyd) (see p.1213) has affected artefacts and hominid remains, of late middle pleistocene age, as well as introducing and mixing various different faunal assemblages. At La Belle-Roche, south of Liège (Belgium), a debris flow has penetrated a low horizontal passage, carrying fauna and artefacts of early middle pleistocene age (Cordy 1980). It seems highly likely that debris flow was, and still is, a major agent in the development of cave sediments, at least in Europe. Such deposits must often have gone unnoticed; examination of the archaeological and palaeontological literature shows that, even when advanced wear or fracture of bones, artefacts and limestone clasts has been recorded well within caves, excavators are still too ready to blame such processes as 'cryoturbation'. As will be seen below, failure to recognise debris flows may have repercussions far beyond the intrinsic sedimentological interest.

The area of the cave of Font de Gaume (Dordogne) known as the "premier diverticule" lies more than sixty metres from the present entrance. For our purposes we will define three groups of deposits, covering Layers A (top) to N.

> Couches A-B-C: (. . .) . . . un sédiment hétérogène, où se trouvent associés des fragments de roche calcaire, des débris de stalactites et des morceaux de plancher stalagmitique. La présence d'amas de sédiment comparable à celui des couches E à J, et celle des vestiges paléontologiques identiques à ceux de la couche M, permettent d'y reconnaître le produit du remaniement des dépôts sous-jacents; il s'y ajoute des apports plus récents, attestés par quelques outils d'une industrie indéterminée du Paléolithique supérieur.
> Le sédiment de diamètre inférieur à 2 mm. est un sable argilo-limoneux brun . . . dont la médiane est voisine de 0,030 mm. La courbe cumulative qui le caractérise présente un tracé aplati et irrégulier, dans lequel on reconnaît le mélange de plusieurs stocks sédimentaires différents . . . ; l'indice d'hétérométrie est de 2. (. . .)
> La base du dépôt est toujours brutale et ondulée; elle tronque la plupart des dépôts sous-jacents, plus ou moins profondément suivant les endroits. (Laville 1975:189)
>
> [Trans.: Layers A-B-C: (. . .) . . . a heterogeneous sediment, in which fragments of limestone, the remains of stalactites and pieces of stalagmitic floor are associated [?]. The presence of accumulations of sediment comparable to that of Layers E to J, and of palaeontological remains identical to those of Layer M, point to the reworking of the underlying deposits; in addition, more recent contributions are attested

by a few tools from an undetermined upper palaeolithic industry.
The sediment below 2mm diameter is a brown . . . clayey silty sand, with a median of c.0.030mm. The characteristic cumulative curve shows a flat and irregular form, in which the mixing of several sedimentary stocks may be recognised . . . ; the 'heterometry index' has a value of 2 [i.e. the sediment is very poorly sorted]. (. . .)
The lower boundary is always abrupt and undulating; the majority of the underlying deposits are truncated, to a greater or lesser depth depending upon the exact location.]

The second group comprises Layers D to L, a sequence some 60cm deep. These are mainly well bedded sand, silt and colloid lenses and beds probably referable, as Laville points out, to flowing water at various levels of competence. The third group includes Layers M and N.

Couche M: Epaisseur maximale: 0,40 m.
Formation hétérogène, où l'on rencontre en très grand nombre des fragments de roche calcaire, des débris de stalactites et des morceaux de plancher stalagmitique, répartis de façon désordonnée dans un sable argilo-limoneux de couleur gris foncé . . . très plastique et collant. L'ensemble présente un aspect chaotique . . . [infra].
Le sédiment de diamètre inférieur à 2 mm. donne une courbe cumulative superposable à celle qui caractérise le sédiment fin de l'ensemble A-B-C . . . .
Couche N: Epaisseur maximale: 0,05 m.
Sédiment présentant les propriétés texturales identiques à celles du matériel fin de la couche M, mais de couleur brun-jaune. Les fragments de roche calcaire et les débris de stalactites y sont absents. (. . .) (ibid., pp.191-2)

[Trans.: Layer M: Maximum thickness: 0.40m.
A heterogeneous formation, in which there are very large numbers of limestone fragments, remains of stalactites and pieces of stalagmitic floor, dispersed in a disorganised manner in a dark grey . . . clayey silty sand which is very plastic and sticky. The whole unit has a chaotic aspect . . . [omission to be quoted separately below].
The sediment below 2mm diameter gives a cumulative curve which may be superimposed [with little or no divergence] upon that which characterises the fine sediment of unit A-B-C . . . .
Layer N: Maximum thickness: 0.05m.
A sediment with identical textural properties to the fine material of Layer M, but of a yellow-brown colour. Fragments of limestone and remains of stalactites are absent. (. . .)]

Now, Laville's descriptions of groups A-B-C and M-N leave little doubt in the present author's mind that these are debris flow deposits. Normally, Laville might have called them solifluction (= gelifluction; see p.360) deposits; it is unclear why he did not, since he was not deterred by the distance from the

entrance and the consequent improbability of a frozen substratum (infra). Laville's own explanation is as follows:

> - le troisième regroupe des dépôts de texture variée (couches M à E), mais le plus souvent caractérisés par une composante argileuse plus importante; ils sont le résultat d'une sédimentation plus calme, où l'on reconnaît des phases d'inondation et de décantation successives;
> - le quatrième enfin (formation A-B-C), correspond en totalité aux remaniements de dépôts antérieurs. (ibid., p.193)
>
> [Trans.: - the third [sedimentary unit; the first and second are not discussed here] groups deposits of varied texture (Layers M to E), but which are usually characterised by a colloid component which is more important [than that of layers deeper than N]; these deposits are the result of calmer [than in layers deeper than N] sedimentation, including successive phases of flooding [apparently gentle] and decantation;
> - finally, the fourth (formation A-B-C) corresponds as a whole to reworking of earlier deposits.]

Note that Layer D is a stalagmitic floor disrupted by Layers A-B-C. Layer N, however, is nowhere mentioned in Laville's interpretation. We are therefore asked to believe that these deposits were formed by calm subaquaeous sedimentation and by some unspecified reworking process. If we wonder what the blocks of limestone and stalagmite are doing in Layer M, Laville has an answer, if not for their presence, at least for their distribution:

> L'ensemble présente un aspect chaotique qui pourrait être le fait du remaniement des dépôts, au fur et à mesure de leur mise en place, par le passage répété des Ours; c'est en effet de ce dépôt que provient la plus grande partie des ossements de cet animal récoltés par F.Prat. (ibid., p.191; excerpt filling the annotated omission in the main passage quoted above)
>
> [Trans.: The whole unit has a chaotic aspect which could be the result of reworking of the deposits, as they were progressively emplaced, by the repeated passage of Bears; indeed, it is from this deposit that F.Prat [excavator and palaeontologist] collected the great majority of the bones of this animal.]

Colourful as this image may be, it is not wholly convincing. As for these bears, Laville's footnote (ibid., p.195) indicates that Prat has suggested that the phylogenetic stage involved might be 'pre-rissian' in date. Since Prat uses a morphometric method of phylogenetic stage identification for bear teeth, involving the treatment of several individual teeth as a grouped sample,

one might, at least in theory, question the validity of his suggestion in this case, on the grounds that at least Layer M may contain teeth from a wide chronological range. Similarly, Laville's comment in this footnote that the supposed age of the bear remains casts doubt upon his own original suggestion of a 'würmian' age for the deposits is just as unacceptable as to say that the palaeontological remains derived from Layer M, but found in Layers A-B-C, indicate that Layers A-B-C are of comparable age to Layer M. That Laville does not make this latter suggestion only shows that his arguments are inconsistent.

Further into the cave of Font de Gaume, somewhere between 60 and 130m from the entrance, in the "galerie principale", more extensive excavations have been carried out (Prat & de Sonneville--Bordes 1969). Of interest here are Layer 3 and Layers 4, 5 and '4-5' (the latter considered by Laville (1975) to be a lateral equivalent of Layer 5 where Layer 4 is missing). Laville suggests that Layers 4 and 5 were emplaced "par ruissellement régulier [by regular stream action]"(ibid., p.198), whilst Layer 3 was emplaced "par ruissellement diffus ou par solifluxion [by diffuse stream action or by gelifluction]"(ibid., p.197). No mention is made of any bedding features that might suggest running water, and no explanation is given of how gelifluction might operate this far inside a cave. Without going into detail, the present author suggests that Layer 3 is almost certainly a debris flow and Layers 4 and 5 are probably also referable to fast and fluid mass movement. In any case, even Laville accepts that these are deposits, including a component derived from outside the cave, which have been emplaced by highly dynamic processes.

Prat states:

. . . dans les stations du Périgord, la présence du Bouquetin (Capra ibex)

> indique vraisemblablement des conditions climatiques assez rudes . . . . (Prat & de Sonneville-Bordes 1969:117)
>
> [Trans.: . . . in the sites of the Perigord, the presence of the ibex probably indicates quite harsh climatic conditions . . . .]

In Layer '4-5', ibex is associated with hedgehog (Erinaceus europaeus) and shrew (Sorex cf. araneus), an association about which Prat makes no comment but which would not appear to be consistent with 'quite harsh climatic conditions'.

De Sonneville-Bordes (ibid.) recognises two flint industries, one "Aurignacien typique" and the other "Périgordien à pointes de Châtelperron"; the former was recovered from Layer 3 and the latter from Layer '4-5'. These identifications are of course based on typology, both with respect to individual 'type fossils' and to the overall character of each 'industry'. Again arguing from typology, she also states, concerning the artefacts of Layer 3:

> Quelques silex se rattacheraient plus facilement à la couche sous-jacente de Périgordien à pointes de Châtelperron, tout au moins les nos. 6, 10 et 11 de la fig.4. (ibid., p.127)
>
> [Trans.: Some flints may be more easily linked with the Lower Perigordian of the underlying layer, at least those shown as nos. 6, 10 and 11 of fig.4.]

In their conclusion, Prat and de Sonneville-Bordes state:

> Les traces d'utilisation (usure, microdenticulations, esquillages) relevées sur divers silex de la couche 3 . . . sont peut-être à mettre en relations avec les gravures qui existent sur les parois de la galerie principale et qui abondent sur celles de la salle des petits bisons. (ibid., p.131)
>
> [Trans.: Use traces (wear, microdenticulations, splintering) noted on various flints from Layer 3 . . . may perhaps be connected with the engravings which exist on the walls of the main gallery and which abound on the walls of the 'chamber of little bisons'.]

Microdenticulations are shown in the figures as irregular notches and 'nibbling', sometimes alternating, and 'splintering' presumably refers to the two "pièces esquillées", which are said to be "des ébauches [rough or incomplete forms]"(ibid., p.125). Given the probable mode of emplacement of the deposit, might not

the anthropogenic nature of these modifications be doubtful? Pieces with strong wear traces (macroscopic wear facets) are most unlikely to be formed in debris flows, however. But the piece chosen for the photograph in their fig.5, presumably because it is the most 'typical', is also drawn as their fig.4, no.11: "Fig.5 - . . . pièces usées de la couche 3; . . . à droite, extrémité de la lame de la fig.4, no.11 . . . .[Fig.5 - . . . worn pieces from Layer 3; . . . on the right, the end of the blade shown in fig.4, no.11 . . .]"(ibid., p.128, caption to fig.5). The other photograph in fig.5 is of a scraper front. As we have seen above, the artefact drawn as fig.4, no.11, 'may be more easily linked with the Lower Perigordian of the underlying layer'.

Therefore, at Font de Gaume, failure to recognise the contextual implications of the deposits has resulted in assumptions concerning the phylogenetic stage and thus age of bear remains; the association of 'cold' and 'warm' faunal indicators; the cultural affinities of groups of flint artefacts; and the function of certain artefacts which would link them with otherwise unassociated cave-wall engravings and thus, by extension, would date the engravings. All this might be the result, or partially the result, of geological processes. Laville was involved in this work before 1969; his help is acknowledged by Prat and de Sonneville-Bordes (1969). Even if his research had not progressed sufficiently at the time of the 1969 publication, why did he make absolutely no comment in 1975?

H. de Lumley-Woodyear published his doctoral thesis, Le Paléolithique inférieur et moyen du Midi méditerranéen dans son cadre géologique, in two parts in 1969 and 1971. In 1974, J.-Cl. Miskovsky published his doctoral thesis, Le Quaternaire du Midi méditerranéen. This is mentioned here because some of

what follows appears verbatim in both texts, without the slightest reference to the other author. The present author is therefore uncertain whom he should credit.

The cave and shelter of La Baume Bonne de Quinson (Basses-Alpes) contains, among other things, a sequence thought to date from the Riss III. These deposits indicate:

> . . . un climat très rigoureux, froid et humide. Tous les galets et outils en grès sont extrêmement gélivés.
> Les dépôts du Riss III comprennent en fait cinq ensembles de couches comblant chacun une rigole de ravinement creusée dans l'ensemble immédiatement antérieur. Les dépôts accumulés dans chacune de ces rigoles ont tour à tour été soliflués; silex et galets furent redressés à la verticale. Il est vraisemblable que la masse du remplissage restait gelée pendant toute l'année. En période estivale, le dégel ne devait affecter que la couche superficielle remplissant la rigole (mollisol) qui se mettait alors en movement dans le sens de la plus grande pente (vers l'extérieur de la grotte) sur un substrat gelé en permanence (pergélisol).
> Entre le dépôt de chacun de ces cinq ensembles, une augmentation de l'humidité favorisait le creusement des rigoles de ravinement.
> (de Lumley-Woodyear 1969:247, and Miskovsky 1974:154)

> [Trans.: [They indicate] a climate which was very rigorous, cold and wet. All the pebbles and tools of sandstone are extremely frost fractured. The deposits of the Riss III in fact comprise five groups of layers, each filling an erosional channel cut into the immediately underlying group. The deposits accumulated in each of these channels were in turn soliflucted; flints and pebbles were stood on end. It is probable that the mass [main part] of the fill remained frozen throughout the year. In the summer, the thaw would only affect [appears to have affected only?] the superficial layer filling the channel (mollisol) which then began to move, according to the greatest slope (towards the cave exterior), over a permanently frozen substratum (pergelisol).
> Between the deposition of each of these five groups, a rise in humidity [precipitation] caused the cutting of the erosional channels.]

The present author considers the above explanation to be almost as unlikely as to be impossible. This whole sequence seems to have accumulated by debris flow.

> Nous avons dégagé dans plusieurs niveaux . . . du Riss supérieur divers empierrements de galets successifs (fig.154). Ils ont été essentiellement installés dans les zones de rigoles qui, en période humide, drainaient forcément les eaux de la caverne et dont le sol devait être à l'état boueux.
> Ces sols empierrés auraient donc été installés par l'homme pour se protéger de l'humidité et vivre sur un terrain plus solide.
> Leurs contours sont nettement délimités (forme ovale ayant le plus souvent environ 5m de long sur 2,50m de large); ils paraissent correspondre à des sols de cabanes. (de Lumley-Woodyear 1969:249)

[Trans.: In several levels of the Upper Riss [Riss III in particular] we have uncovered various successive 'pavements' of pebbles (fig.154). They have been installed essentially in the areas with channels; during humid periods, these channels of course acted as drains for the cave waters and their floors would have been muddy.
These paved floors would therefore have been installed by man in order to protect himself from the humidity and to live on more solid ground. Their boundaries are clearly defined (oval form, usually about 5m long by 2.5m wide); they would appear to be cabin [hut] floors.]

In de Lumley's fig.154, the stone concentrations are only shown as occurring in the bottom of channels; some of them are only 50cm wide. Photographs of one of these floors are available; two versions have been seen by the present author, one in Miskovsky (1974, p.152, photo 6) and the other in de Lumley and Boone (1976a, p.642, Fig.22). Debris flows very commonly have basal deposits such as these. In any case, has anyone stopped to consider why these people should have been so perverse as to sit in the middle of a stream when they could have camped on the drier, and safer, banks above, thus saving themselves the considerable trouble of collecting thousands of pebbles? At least Laville's bears kept moving and did not try to set up house.

Les dépôts datés du Riss contiennent une industrie . . . assez homogène de la base vers le sommet du remplissage. Nous avons pris comme type celle de la couche D (Riss III), l'une des plus riches.
(de Lumley-Woodyear 1969:250)

[Trans.: The deposits dating from the Riss contain an industry . . . which is quite homogeneous from the base to the top of the sequence. We have selected as the type [collection] that [presumably industry] from Layer D (Riss III), one of the richest.]

Not only is the 'industry' homogeneous within the <u>whole</u> Riss sequence, it is spatially homogeneous as well:

Il est ici évident que les différents outils étaient répartis à peu près uniformément à travers tout le gisement; . . . . (ibid., p.251)

[Trans.: It is clear from this [the comparison of various cumulative graphs of tool types from different parts of the site] that the different tools were distributed almost uniformly across the whole site; . . . .]

De Lumley takes nine pages to detail the typology of the 'industry' of Layer D. It is no surprise that more than 40% of

the 'tools' formally involve notches and that the whole is labelled 'Tayacian'. One wonders if it was wise to use this material as the 'type', especially when one reads:

> A la fin du Riss une augmentation notable de l'humidité a provoqué un important ravinement des niveaux sous-jacents. (de Lumley-Woodyear 1969:247, and Miskovsky 1974:154)
>
> [Trans.: At the end of the Riss a marked rise in humidity caused major channelling of the underlying levels [process unspecified].]

Miskovsky (and only Miskovsky) adds, immediately after the statement quoted above:

> Ceci explique l'extraordinaire concentration de silex dans la couche D. (1974:154)
>
> [Trans.: This explains the extraordinary concentration of flints in Layer D.]

Could not such reworking account for the featureless distribution of artefacts, the high proportion of 'notches' (cf. damage) and thus the very identification of the archaeological culture involved? The only trait which really sets the Tayacian apart from other industries of similar age is a very high proportion of notches and associations of notches ('denticulates', 'Tayac points', etc.).

Further examples of the difficulties concerning the interpretation of artefacts and fauna recovered from debris flows are given below in Part II (see especially Pontnewydd, p.1208).

9.4.3. Avalanche

The word 'avalanche' is used here as a generic term to cover mobile rockslides, landslides and true avalanches (associated with snow and ice)(cf. Statham 1977; Brunsden 1979; Embleton 1979). The mechanisms involved are probably similar in the whole range, from pure rock to pure snow/ice. Naturally, a snow/ice avalanche

must have contained significant rock debris if a fossil deposit has remained.

Brunsden notes:

> The plan form of the debris is similar to that of a lava flow or glacier. If unrestricted it extends as a wide lobe across a flat plain. If confined to a valley it extends down valley but still possesses a lobate margin. Where a slide descends into a narrow valley floor it commonly blocks the valley, forms lakes, and may run up the opposite side . . . . A typical feature is the ability of the slide debris to move over or around major obstacles. The area covered is always large and the volume is measured in millions of cubic metres.
> The longitudinal profile of the debris lobe shows a gently sloping surface with very low relief, ... There is sometimes a trough between the head [proximal end, infra] of the slide and the base of the scar. The surface is hummocky, may include ponds and enclosed depressions, and is marked by both longitudinal and transverse ridges produced by shear between debris streams or as pressure ridges. The edges are sharp, marked by levées and with a pronounced distal rim.
> The material is generally angular but remarkably unbruised, and heterogeneous in grain size with a tendency for fine material to concentrate in the lower layers, coarser material appearing on the surface. Very large clasts, often measuring several metres in equivalent diameter, may be transported. Some particles and surfaces may be striated but a common form is a jig-saw effect where blocks have been fractured but not segregated. The debris shows considerable ordering, the original stratigraphic succession being preserved, and is very porous, especially in recent debris. (1979:159-60)

Such phenomena result from the collapse of near vertical faces. Thus the term "head" refers not to the distal portion, as in most mass movement, but to the proximal end which was once the 'highest' part of the source (cf. 'headwall'). Although Brunsden's requirement of a huge scale is questionable (cf. Rapp 1960), it is clear that these must be large scale events in order to preserve a coherent pattern. Note that they are not usually erosive as such (due to basal sliding and air-cushion gliding), although they will incorporate unstable deposits and may erode bedrock if confined to gullies during the descent. Another diagnostic feature is their asymmetry if they have descended onto a steep valley floor at right angles; on the up-valley side the debris will be very thick and will thin rapidlt down-valley, even if the lobe as a whole shows no signs of having been diverted

down-valley.

Deposits of this type have not yet been recognised in or even near cave mouths. However, it should be noted that avalanching is an important periglacial process, even in areas of comparatively low relief. Avalanches are the logical end member of the speed range of mass movement; velocities of 400km/hr have been estimated during some very large events.

10. MECHANICAL PROCESSES ASSOCIATED WITH WATER MOVEMENT

10.1. Wash Processes

> Rain falling upon the land surface is split into two parts, that which flows across the surface and that infiltrating and flowing through the soil. The surface-flow fraction forms an intermittent water-transport system whose important property is that there is no concentration of flow into specific flow-lines or channels through a long period of time. That is to say, flowing water moves as a continuous smooth sheet across the whole slope, or it gathers into small channels or rills whose positions shift from time to time and therefore affect the whole slope over long periods. Its erosive effect is often called wash transport. [...] Run-off may be generated in two circumstances, either when rainfall intensity ... exceeds the rate at which water can percolate vertically into the soil ..., or when it exceeds the rate at which water drains laterally through the soil. The former may be termed infiltration excess overland flow and the latter saturation overland flow. (Statham 1977:124-5)

The distinguishing features of wash processes are therefore that true water transport (and not just water-lubricated movement) is involved, and that flow is not concentrated into stable channels. However, in steep limestone terrain, wash may become highly concentrated, both along a whole series of limestone surface forms (karren; cf. Bögli 1960) and along any type of re-entrant form in a hillside or cliff (gullies, chutes, etc.). Thus, channels may be relatively long-lived, even though water flow is only intermittent. Wash on limestone that does not have a significant sediment cover will be dominantly 'infiltration excess overland flow', and it often has an almost immediate response rate to rainfall. Under a wide range of climates, limestone with a thin sediment cover, the mineral fraction of which is composed of limestone fragments and local residual

fine material, will also often have rapid wash, since such thin cover usually contains extremely high proportions of fibrous organic material which is radically hydrophobic when in an initial dry state. Luckily, most limestones, although they are impervious on a small scale, are able to absorb huge quantities of water via fissures of various dimensions, so that the proportion of meteoric water removed by overland flow is relatively small. Nevertheless, impressive 'torrents' may often develop during storms. The water reaching a cave through fissures, as more or less persistent drips and sheetflow, is also capable of transporting particulate matter, derived either from the limestone or from the superficial sediments through which it has passed. This is not, strictly speaking, a wash process but, since the water emerges into a sub-aerial context in the cave before depositing its load, the sedimentary result may be very similar. It should not be forgotten that, in addition to rainfall, water may also be provided by melting ice and snow.

It is clear that, in the cave context, there will usually be two quite different zones of wash activity. First, the exterior landscape will have varying proportions and textures of bare rock, sediment and vegetation, all of which will be subject to a wide range of water input processes (rainfall, ice- and snow-melt) and water movement processes (rainsplash, sheetflow, rillflow, etc.). In most situations, slope angles will be relatively high, at least in proximity to (immediately above) a cave or shelter. Second, the cave or shelter floor itself will usually be composed of unvegetated sediment (although rock sills sometimes occur at cave mouths), which will not be subject to rainfall (given a sufficient overhang), but which may be subject to localised drips and sheetflow. Caves that have accumulated coherent sequences

of deposits of interest to the quaternarist will normally show relatively low bedding angles, except in the special case of steep talus cones which develop below restricted entrances. As a generalisation, these two zones may be characterised as the zone of sediment erosion and transport (the exterior) and the zone of sediment deposition (the cave). As entrance slopes in a cave increase, the zone of erosion and transport may penetrate further into the cave. Normally, however, most material will be deposited quite near the entrance, and only material carried by water arriving through fissures will reach the deeper areas of the cave. The topics of interest are therefore: the processes of erosion and transport; the available sediments; the 'catchment' of a cave entrance; the nature of the 'break of slope' between the exterior and the cave; and the processes of deposition.

The material mobilised by wash is a function, on the one hand, of the nature of the water input processes and, on the other, of the nature and erodibility of the local sediments. The important wash processes are rainsplash, sheet wash and rill wash (and even more concentrated flows).

Rainsplash is an impact process. On level ground, rainsplash tends to cause random reorganisation of particles, due to pushing, undermining and rebound. On sloping ground, the trajectory of raindrops, now striking the surface at greater angles, sends increasing proportions of particles downslope as the slope increases, an effect which is enhanced by the downslope component of gravity. Also, longer rebound trajectories are possible downslope. Moseley (1973) has demonstrated these effects experimentally, as well as the fact that (other things being equal) total transport increases with slope due to increasingly tangential impact trajectories with more efficient erosion. The

energy of a falling raindrop is a function of its size and its fall speed. Since fall speed increases with drop size, momentum increases very rapidly with rising rainfall intensity. According to figures given by Pettersen (1958), the difference between 'fine rain' and 'storm rain' may be approximately expressed as a tenfold increase in drop size, a threefold increase in fall speed, with an increase in momentum by a factor of over 3000. It is therefore obvious that the effects of rainsplash are proportional to the intensity and frequency of rainstorms, with total rainfall playing a less important role.

Sheet wash is generated by rainfall and melting of snow. The melting of ground ice produces water that is intimately associated with the sediment; either mass movement will occur or water will drain, if it reaches the surface downslope, in a more or less channelled fashion. On irregular ground, with impervious sediment or rock near the surface, this water may, of couse, feed sheet wash lower down the slope. Sediments already saturated by ice-melt will raise the erosive power of additional water supplied by rainfall or snow-melt (cf. 'saturation overland flow', supra). The ability of flowing water to erode sediment is closely linked with the nature of turbulence, which in turn is a function of flow depth, flow speed and the roughness of the 'bed'. Smooth, thin sheet wash will have little effect but, as the sheet becomes thicker or as rill wash starts to develop, the erosive power and competence of the flow increase.

Rainsplash, sheet wash and rill wash are linked in a dynamic equilibrium. Since erosion requires more energy than transport, thin sheet wash, which would be ineffectual on its own, may transport the material eroded by rainsplash. However, as the sheet wash thickens, it buffers the effect of rainsplash, thus

reducing overall erosion. On a rilled surface, rainsplash may be active on the 'interfluves', supplying material to the rills, which are themselves developing more turbulence and thus erosive and transport power.

It has often been noticed that wash erosion decreases after a time, even when the flow regime remains essentially constant. This has been variously explained as due to rainbeat (raising surface compaction), reduction of surface irregularity by the sedimentation of fines between coarser particles (reducing turbulence), or selective erosion of finer particles (causing armouring of surfaces by coarser residuals). When the same surface is allowed to dry and is then subjected to further rainfall, erosion is again active for a while. This indicates that drying loosens the surface particles, ready for another cycle; frost may have a similar effect.

Vegetation plays an important role in the control of wash erosion. Foliage intercepts rainfall and reduces rainsplash, it intercepts wash and slows the flow, and the root networks increase permeability, bind the sediment and reduce erosion. However, the effect is not absolutely straightforward. Under closed woodland, for instance, leaves high on trees may concentrate water (even from 'fine rain') into very large drops at their tips, and the resulting impact at ground level (where there may be little cover, due to low light intensities) may still be effective. Also, much water is concentrated down trunks and may cause rilling at the base. Statham (1977) notes a general relationship between vegetation cover and precipitation: in very arid environments there will be much vulnerable bare ground but little precipitation, whilst in wet environments there will be plenty of water but much more efficient protection by vegetation.

This model would result in the most effective wash occurring on moderately watered and vegetated areas. However, the present author doubts whether this effect will be sufficiently general to be useful as a diagnostic principle, since local topographic factors are so important and highly fluctuating water availability is known to cause the greatest erosion. In limestone areas, vegetation will often be a function of the angle and stability of slopes and not so much of annual rainfall. It would be dangerous to suggest that maximum wash should occur at 'medium' levels of vegetation, especially since this would entail the temptation to interpret pleistocene wash deposits as occurring in 'better vegetated' and thus 'warmer' environments. The complexity underlying wash erosion is well illustrated by Bovis and Thorn (1981) in their discussion of soil loss in an alpine zone in Colorado. Slopes range from 5-30° and soils are mostly non-cohesive. Tundra meadow, with a herbaceous cover, suffers surface lowering in the order of 0.01mm/y. Dry tundra, with 50% bare ground, suffers lowering in the order of 0.1mm/y. These results are predictable from general theory. However, near snow patches, even on relatively low slopes (nivation hollows), lowering is in the order of 1mm/y (not including mass movement), even though such areas are often vegetated. In this study, nivation hollows, representing only 3% of the surface area, produced 50% of the eroded sediment caught in traps. Similar observations have been made in cool and cold environments by many other researchers.

The nature of the sediment controls erodibility. De Ploey and Mücher (1981) have discussed the effects of rainwash on loamy Belgian soils. The most resistant soils have high clay and organic contents, a larger variation in pore size, carbonate grains,

and sepic plasma fabrics (oriented silt and colloid domaines). Thus, these soils are generally more structured and more cohesive. Loose sediments are usually much more erodible, with well sorted material of (fine) sand size being most vulnerable. However, Poesen (1981) has shown that the initial water content, before rainfall, is also very important when considering loose sediment; the drier the sediment, the greater is its erodibility. Recent work by Summer (1982), in the Rocky Mountains, suggests that erodibility may not in fact be a particularly predictable property. She was unable to find strong correlation between field measurements of erodibility and a range of physical properties (eg. clay, organics, etc.) later measured in the laboratory. The surface morphology of deposits is important and, far from protecting underlying sediments, coarse debris and even sparse vegetation may sometimes increase erosion by 'channelling' flow and increasing turbulence (cf. de Ploey et al. 1976). It should be noted that the maximum grain size that can be moved by rainsplash and sheet wash on steeper slopes has been reported by many observers to be frequently in excess of 1cm.

Judging from published observations in a range of environments, it would appear that wash processes often mobilise significant amounts of sediment. This would seem to be particularly true of areas with pronounced relief (cf. Caine 1976; Rapp & Strömquist 1976; Gardiner 1979). Climate is certainly a factor, but wash processes in general are not indicative of particular climatic types. Thus, French (1976) stresses the importance of slope wash in periglacial environments, whilst the present author has noted the accumulation of over 50cm of predominantly wash deposits in the mouth of Pixie's Hole (Devon) in a little over a century since quarrying at the site was stopped (the source slopes

for the 'colluvium' were in no way disturbed by the quarrying).

Given that sediment is moving on local slopes, it will be the geomorphological relationship between a cave and its surroundings which will control how much sediment reaches the cave. The concept of 'catchment' is very simple, yet so often overlooked. If enormous quantities of material are moving due to wash, a given cave may still not receive significant amounts of material if insufficient slopes converge on the entrance. On the other hand, a site, such as a fossil influent cave in the bottom of a doline, may receive huge amounts of material, even when slope wash is not a particularly active process in the environment as a whole.

Although much work has been carried out upon the nature of erosion by wash processes, very little is known about the sediment bodies resulting from the redeposition of the eroded material - studies are simply couched in terms of 'soil loss'. Since mobilisation by wash is often a comparatively slow process, clear depositional structures may not develop where other types of sediment are also accumulating. Similarly, if wash is reaching a cave entrance deposit mainly from a rockface above, the last stage of sediment input may be by free-fall over a relatively wide area. It seems probable that large amounts of the finer component of many cave sediments may be referable to wash, although structural proofs in given instances will often be lacking.

Common arguments used to suggest sediment input by wash fall into two groups.

First, the identification of allochthonous material, especially in the sand grade, is often said to imply wash processes, as long as evidence for mass movement is absent. Thus, Laville

(1975) uses the presence of ferruginous quartz grains, derived from sand bodies ('Sidérolithique', Upper Cretaceous) perched on the limestone plateaux above his sites, as an indicator of wash. Rounded shiny quartz grains (Fr. émousés-luisants) are quite common in water-laid deposits in caves, so that Miskovsky (1972) uses their proportion in apparently unstructured deposits to indicate wash input derived from within the karstic system of l'Hortus (Hérault); no explanation is given for the fact that the cave passage which lies behind the main entrance sequence, as well as all the sediments in this passage, slope down inwards. Miskovsky (1974) also interprets the rounded shiny grains from the relatively open site of Gramari (Vaucluse) as indicating a humid climate. Since this site is on a point bar of the Nesque, it might be supposed that the presence of these grains might be more related to the hydrography and catchment of the stream than to variations in climate. Quartz grain morphology is a complex subject and the present author would hesitate before equating the proportion of rounded shiny grains with the degree of wash activity. Miskovsky (1972) also uses rising proportions of total heavy minerals as an indicator of humidity (wash?). It is not clear how this proportion is calculated; the 'total sand', of which the heavy minerals are a proportion, could be either whole sand (including carbonates) or only the acid-insoluble residue. In the former case, heavy mineral content merely reflects carbonate content and, in the latter, there is no obvious mechanism to explain the enrichment in heavy minerals, since the mineral types are identical throughout the sequence at l'Hortus and there is no sign of different source stocks (according to Miskovsky).

The second type of argument often used to suggest wash input concerns the particle size distribution, especially of the

non-carbonate sands, which is said to indicate energy of transport. The better the sorting and the coarser the median of the sands, the greater is the wash input suggested by Laville (1975). Miskovsky (1972) seems to place more importance on the presence of fine sands, although the reason for this is never clearly explained. This sort of simplistic equation between texture and process is bound to be misleading. Butzer (1981) has shown that more credible information concerning wash and other input processes can be extracted from textural data, but only with a much more detailed approach (cf. section 3.3.).

Finally, Miskovsky (1969) has combined compositional and textural data from Le Lazaret (Alpes-Maritimes) in order to indicate the 'intensity of wash'. In his Fig.23 (ibid., p.38), he shows three graphs plotted against the layer sequence: the percentage of fine sand; the percentage of rounded shiny quartz grains; and the 'index of flattening' of small 'karstic' (non-carbonate) pebbles. The first two properties are percentages and the last is an index, the units of which are arbitrarily plotted as paper intervals equal to 5% on the other two graphs. A fourth graph, "une courbe moyenne [an average curve]" (ibid.), is then constructed which is a simple arithmetic average of absolute distance of points on the three saw-toothed lines from the abscissa, that is, the 'intensity of wash' is purely a function of the arbitrary choice of the paper intervals of the scales (cf. section 6.3.). Fig.11 (p.36), Fig.19 (p.38) and Fig.21 (p.38) of this same publication also purport to show fine sand, rounded shiny quartz grains and flattening of small 'karstic' pebbles, for the same layer sequence; each one of these separate figures gives different results from those shown in Fig.23. Assuming that these different results refer to a different area

of the cave, one wonders whether Miskovsky took lateral variation into account when evaluating the 'intensity of wash' and thus the wetness of the climate. Also, why was the flattening index rather than the proportion (ibid., Fig.20, p.38) of 'karstic' pebbles used in the calculation of the 'average' curve, and why were other components (eg. colloids, silts, heavy minerals), all said to be allochthonous and all figured in graphs or reported in tables, not included in the calculation? A grand 'average' graph, using all seven of the parameters mentioned above, looks rather different from Miskovsky's curve, but it is no more indicative of the 'intensity of wash' than was the original.

Structures that appear to be due wholly or partly to wash processes are usually some form of cyclical bedding. These may be divided into two groups: those which are due to fluctuations in the wash processes themselves and those which are due to alternations between wash and other processes. One further set of particulate structures, which may be referred to collectively as 'mud clasts', may sometimes be associated with wash.

Belgian and Dutch researchers have recently carried out many experiments on the redeposition of loess by wash processes (cf. Mücher & de Ploey 1977; Mücher, de Ploey & Savat 1981), the results of which have been successfully applied to pleistocene open-air sequences (cf. Mücher & Vreeken 1981). The source material for the experiments was 'clean' loess, with most particles in the range 10-50 microns. This restricted source material obviously affected the results (infra), so that care must be taken in the extraction of generalisations. Nevertheless, some of the results are extremely interesting and similar experiments with a range of source sediments would certainly prove most informative. When water input was restricted to surface flow (analogous to meltwater

flow or the after-flow of rainwash), the silts formed very clear laminations, with good sorting of particle sizes and mineral species. These laminations were up to 0.2mm thick during normal flow, but reached 1.5mm for the heavily loaded waning flow after water input had ceased. It was found that concentration of flow into rills inhibited the formation of laminations. Rainsplash also inhibited good bedding but, as long as sheet wash was also involved, some less distinct laminations were still formed. During flow without splash, the finer silt fractions were preferentially deposited, whilst turbulent rainwash with splash deposited the whole range of particle sizes. It is not clear whether this last feature has any relationship with sorting at the initial mobilisation stage, although this would seem likely to the present author. Laminations were observed in thin section under an optical microscope; it is not stated whether they were macroscopically visible but, given the homogeneity of the source, it seems unlikely that they were.

These experiments have important implications for cave sedimentology. First, it seems reasonable to suggest that the observation that splash inhibits the formation of laminations in loessic material may be generalised to cover most fine sediments. Given that a sediment will indeed form laminations in the absence of splash, we might expect more massive deposits to grade into laminated deposits, either as we move into the shelter of the rock overhang or as we move away from zones of significant drip or rilling within the cave. On the other hand, if water is mainly available as melt, laminations may be present in the open as well, with more massive deposits restricted to drip and rill zones. The question is, then, under what circumstances might a range of sediment types be expected to

form laminations, given the absence of splash.

The use of clean silt in the above experiments meant that regular deposition would only occur on very low slopes (<2°), implying low flow velocities (not reported). This is predictable from standard Hjulström (1939) and Sundborg (1967) diagrams, since clean silts at c.30 microns will certainly be eroded at velocities above 2km/h (56cm/sec). We may also suggest that silts will show the most persistent laminations of any sediment deposited by wash, because this size grade is most sensitive to the density of the suspension load. A slight reduction in the load (deposition) means that flow velocity must fall significantly before more sediment of the same size is deposited. Thus, on a long stretch of low angle slope, with slowly waning flow velocity (due to water infilration into the deposits, or to a further decrease in slope), a generally constant lamination might be formed. As the sediment becomes coarser, more and more material will be moved by traction and it will not be very sensitive to the total amount of sediment in movement. Most coarser sand, therefore, should be 'dumped' rather quickly as the slope declines, with the possibility of the formation of 'microdeltas' at strong concave-up breaks of slope. Only when slope and flow velocity are reasonably constant (eg. in mature stream flow) will sand laminations be extensive. The finest silts and the colloids, on the other hand, will only be deposited if water flow almost ceases or if they are flocculated by reaching a more alkaline environment. Optimum conditions for the formation of the finest structures would therefore seem to be very low flow velocities (either in terms of low slope, only small amounts of water, or very rapid infiltration), perhaps on irregular and damp (and thus often alkaline) surfaces. Laterally extensive bedding will almost

certainly be restricted to gently flowing or still, subaqueous conditions (ponding). Once colloids have settled and compacted over only a few hours, very much higher energy levels will be necessary to erode them, levels which are only likely to occur during rilling on higher slopes (leaving aside processes of radically different types). Therefore, on steeper slopes, with small or medium scale irregularities (cf. limestone fragments of various sizes), we might expect rather small scale braided patterns to occur, with rapid lateral and vertical shifts in sediment texture.

These theoretical considerations suggest that wash deposits are a prime candidate for a facies model approach. The real sedimentary results of wash will obviously be affected by many local factors (such as overall slope, microtopography and availability of different sediments) as well as the rainfall and meltwater regimes. For instance, the author has observed laminations forming at an open site below a limestone outcrop in Velvet Bottom (Mendip, ST 48525458). Quarrying (Black Rock Quarry, now disused) has breached an ancient cave passage choked with well bedded and ochreous (yellow) sands and silts, material which is acting as a source for wash deposits on the c.15m slope below. Sediment moving in the environment in general tends to be even finer, consisting mainly of red fine silts and colloids, including organics. These two components are associated in different ways. On steep slopes (up to c.40°), discontinuous units can be seen that comprise alternations between comparatively coarse yellow bands and clayey, slightly fissile (thin and extremely localised lenticular bedding?) redder bands, with sharp boundaries between the two. This is interpreted as normal slopewash (redder bands), interrupted periodically by major input from the ancient cave

sediments (yellow bands), probably controlled by slumping and partial liquidisation at the headwall. In depressions (and even footprints!) on the slope, graded laminations are present, passing from yellower and coarser at the base to redder and finer at the top. This is interpreted as settling out of a mixed sediment in a 'puddle' after rain. At the foot of the slope, at angles generally below 10°, an 'apron' is forming, composed of more or less homogeneous red-yellow sediments, with only minor traces of macroscopic laminations. Here, the two stocks have been blended and microscopic examination would be necessary to investigate the structure. Were these three different expressions of the same environment to be examined separately, each might be interpreted as being due to a radically different 'climate', especially if the two source stocks were not recognised.

Reported observations of possibly wash-induced structures in pleistocene cave deposits usually involve macroscopic laminations. Laminations are by no means always indicative of wash and, even leaving aside more continuous deposition by water (in streams and permanent pools), they may form due to gentle overbank flooding (cf. parts of Layer I at Le Moustier, Dordogne; Laville 1975:171 and 183) or as sub-surface lags at the base of suffosed coarse deposits (cf. p.444). Post-depositional formation is also a possibility (cf. the 'banded fabric' reported by Goldberg, p.269 of the present text). It is therefore clear that more information than the mere presence of laminations is required to distinguish wash deposits.

Some of the observations by Butzer in the Cueva Morin (cf. section 3.3.) are of interest here. Layer 10 is described as "silt loam, with basal or intermediate lenticles of laminated loam"(1981:146). The relatively coarse material of the lenticles

is of Butzer's textural Class 2, said to indicate fluctuating but quite high runoff. Layer 2, however, appears to contain more continuous laminations and is of textural Class 5, said to indicate subdued runoff with mainly suspended load transport. These textural classes represent insoluble residues and not the whole sediments; the present author makes the assumption that, in these two cases, the residues and the whole sediments have similar textural properties. Certainly, these combinations of textural and structural properties are exactly what one would expect from theory.

At Le Lazaret, the description of bedding features, including laminations, and deposit geometry given by de Lumley and Tavoso (1969) provides a very much more convincing argument for wash than do the speciously detailed, textural and compositional arguments of Miskovsky (supra). It is also noteworthy that land mollusca, including taxa that are not troglodític and which are too small to have been collected by man (Barrière 1969), were concentrated in the zones of laminated sediment. Fedoroff (1969) describes thin sections of some of these deposits but, most unfortunately, these are not figured. In the zones without clear laminations, fine sand is irregularly distributed in an oriented plasma with high porosity. This material is crossed, at irregular intervals, by laminations, up to a 'few' millimetres thick, with well sorted, medium and coarse silt in a plasma of very low porosity. Fedoroff makes no comment about these two fabrics, but it seems possible that the unlaminated material (if it has not been disturbed or modified by human agencies) may have been deposited as irregular aggregates rather than as individual particles.

Finally, the small scale features (microdeltas, 'puddle'

deposits, braided rills, 'granule' lags, cross-slope organic arcs, oriented microfaunal remains, drip craters and drip splash lags, etc.), that are so obvious on the contemporary surfaces of many irregular cave deposits (cf. the present entrances at Aveline's Hole, Goatchurch Cavern or Slitter Hole, all on Mendip), will probably only become apparent in buried deposits after very careful 'microsectioning', backed up by microscopic examination at a wide range of magnification.

The structures most commonly cited as being due to episodic wash, alternating with input by other processes, are grèzes litées. This subject has already been mentioned in section 8.3. in the context of cryoclastis, the process thought to be responsible for the initial production of most of the sediment. To the present author's knowledge, no detailed eye-witness account of the formation of grèzes litées has ever been published. Nevertheless, modern occurrences are usually associated with snow banks. Wash may either be supranival, with material being moved onto the snow from the talus slopes above (cf. Jahn 1960), or subnival, with material being moved and sorted during the thaw (cf. Pissart 1967). Reports of structure within the finer unit of the fine-coarse couplets are rare. This may be due to the possibility that, where even the finer units are clast-supported, the principal role of wash is to remove fines from the upper part of coarser units, perhaps with some 'illuvial' concentrations at depth, rather than than to deposit fines on free surfaces (cf. Guillien & Lautridou 1974). In any case, typical grèzes litées should pass laterally into finer wash deposits at the base of slopes, although the present author is not aware of any published description of such an arrangement. Sometimes the finer units can be recognised as being due to aeolian

deposition (cf. Harris 1975), but Lautridou (1971) would prefer to distinguish such deposits as pseudo-grèzes. Wasson (1979), discussing modern stratified slope deposits in the Hindu Kush, notes sequences of thinly laminated silts in units up to 5cm thick and bedded at angles up to 30°, openwork scree, clast-supported scree and extremely viscous sheet mass movement deposits that have sometimes moved over openwork scree without penetration. In pleistocene "grèzes litées" in Tasmania, Wasson (1977) has even identified cut-and-fill structures and cross-bedding. It is clear that there is a continuum between typical grèzes litées and what might better be called 'alluvial fans'. The contribution of wash processes ought therefore to be demonstrated with structural evidence from individual units. As a cautionary note, Wasson (1979) also cites high angle deposits, at sea level in New Zealand and at only 200m O.D. in the Philippines, that have both openwork and clast-supported beds looking very like grèzes litées. These deposits have no obvious link with either wash or cryoclastis, and there is certainly no possibility of snowmelt; they seem rather to be due to dry 'avanching' (grain flow) on slopes close to the angle of repose, with a constant input of sediment from chutes.

Mud clasts are aggregates of fine sediment that have been redeposited after erosion of the original source, the core of a clast still showing the fabric of that source. Individual aggregates may be from sand to fine gravel size, and exceptionally up to boulder size. A particularly impressive instance of such structures is provided by the pleistocene "pellet sands" of Central Europe (Kukla 1975), deposits of which may reach thicknesses of several metres. These are well bedded slope deposits, at various angles, which are composed of sand sized aggregates derived from

the erosion of loamy soils. Kukla reports the formation of such pellets, with aggregates up to 8mm in diameter, observed after torrential rain following a long dry period in Czechoslovakia. The pleistocene occurrences are often underlain by "markers" (aeolian dust deposits), which may be responsible for choking vegetation and thus encouraging the formation of pellet sands, and which themselves show some bedding due to reworking by wash. True pellet sands are probably dependent upon a particular combination of source, topography and meteorology, but the present author has observed relatively high concentrations of mud clasts from limestone terrain south-east of Liège (Belgium). Here, rock outcrops are capped by loamy deposits, probably derived from loess, which are being eroded near the heads of steep gullies. At one particular gully/shelter site, Le Bois Thibaut, loose samples (kindly made available by M. Dewez) were examined from a sequence going back at least to the Bronze Age. Mud clasts were common and, although it is difficult to suggest a lower size limit (the samples had dried, forming artificial aggregates), spherical units of up to 2cm diameter were present. These consisted of a dense silty core, apparently without structure, covered by a thin coat of very finely laminated clayier material. Most units showed a gradual colour change, even within the silty core, darkening outwards from the centre. Some units were partially 'armoured' with sand grains, woody debris and fragments of microfaunal bone and rodent teeth. The author has noted similar structures, but never in such quantities, from various gully sites on Mendip and in the Torbryan Valley (South Devon), both areas where loamy soils, and even some loessic derivatives, are common.

Mud clasts, as such, are not diagnostic of wash and they may form wherever there is a suitable (cohesive) source, especially

if the erosion form involves vertical banks cut into a fine deposit. Clasts (cf. 'mud balls', 'armoured mud balls') have been reported from mountain streams, alluvial fans, glacial outwash, wadi environments, tidal flats and high energy beaches. Significant accumulations of sand sized aggregates, often concentrated by wind rather than water, are quite common in arid and semi-arid environments (cf. Price 1963). Aggregates also occur in mass movement deposits (cf. 'clay clumps', p.349; 'lithorelics', p.380), and in immature alluvial deposits in caves (cf. p.436).

10.2. Permanent Freshwater Bodies

The deposits discussed in this section are associated with more or less permanent freshwater bodies. Exterior lakes and streams may sometimes encroach upon rockfaces, forming, deepening and/or depositing sediments in rockshelters, either at normal flow stage or at flood (cf. Pech-de-l'Azé II, le Moustier, la Madeleine and many other shelters in the Dordogne; Laville 1975). Streams, either influent or effluent, may have run through a cave passage being investigated. Ponding often takes place, producing deposits of various dimensions, similar to external lacustrine sequences. Finally, true phreatic sedimentation may occur in cavities completely filled with water.

There is a continuous gradation between the wash processes, discussed in the last section, and the processes associated with more permanent water bodies. For instance, Frank (1971) describes a 4m thick unit at the Koonalda Cave (South Australia) that consists mostly of water-laid sediments. Frank interprets these deposits as being due to an intermittent stream which was ponded

against a wall of the cave, in a basin perhaps 15m across. There was no coherent throughflow and the water was lost by percolation through the underlying sediments and by evaporation. Much of the sediment consists of graded sandy lenses, up to 30cm thick and with length-width ratios of 50-100:1, and laminated decantation deposits. The latter include occasional surfaces showing desiccation cracks which are interpreted as subaerial phenomena. To one side of the pool, there are cut-and-fill structures, indicating more active flow events. Comparison of thin sections of these deposits with modern exterior soil show quartz grains, with very similar size, sorting, morphology and argillans (clay coatings). Above this unit is another sequence, nearly 2m thick, consisting of interbedded limestone breakdown and fine water-laid sediments. Radiocarbon dates from various levels suggest that deposition of the two units (c. 6m in all) was quite rapid, certainly covering a period of no more than one or two millennia. Obviously, the beds of a sequence such as this must be treated individually and there would be no advantage in lumping everything into an arbitrary 'fluvial', 'wash' or 'ponding' category. Interestingly, Frank attributes no climatic significance to the variation in these sediments and he points out that, since the stream would have had to cross at least 100m of aggrading rock breakdown deposits to reach the pond area, its course would be subject to periodic and more or less random diversion to other areas of what is a large and complex cave system.

Fluvial and lacustrine deposits in the open have been subject to intensive study for over a century and the resulting suites of textural and structural properties are well known (cf. Reineck & Singh 1980). Therefore, a detailed discussion of

such deposits is not called for here. Rather, an attempt will be made to identify some of the peculiarities of alluvial sedimentation in caves.

First, it will be necessary to examine the general sedimentary effects of flowing water. Large bodies of well structured sediment are often associated with rivers and lakes, and observed rates of deposition may be very high. It might therefore be assumed that the main role of these water bodies is depositional. This is very far from the truth: the role of rivers is dominantly erosive. Except in the special cases of closed drainage basins and areas of subsidence, lakes are relatively temporary irregularities that sooner or later will be removed in the general pattern of downcutting. The main reason why fluvial and lacustrine deposits survive is the constant lateral shifting of channels, the zones where most of the erosive power is concentrated. This may be expressed on a large scale, as in the survival of ancient terraces perched high on hillsides, or on a small scale, as in point-bar or abandoned meander sequences. Alluvial deposits may look impressive but they are constructed at the expense of the whole landscape, which suffers a general surface lowering representing many times the volume of sediment that is actually stored for any significant time in the valley bottom.

If we turn to caves, the situation is quite different. Except in very wide passages and larger chambers, there is little scope for a stream to shift its channel laterally. This means that, for the duration of water flow in a given passage, the sediments must suffer the full brunt of any erosive potential developed by the stream. Generally speaking, the time required to develop a new passage is orders of magnitude greater than the

periodicity of fluctuations in water and sediment input. Indeed, the passages of many caves were formed in the phreatic zone, long before streams as such developed and often long before the deposition of any presently extant sediments. If a passage becomes increasingly choked with sediment or there is a relatively massive input of flood water, either the passage sediments will be flushed (cf. section 1.4.) or the flow (or part of the flow) will move to another available passage. Depending upon the specific circumstances, a new route may be above, below or to the side of the original passage. In any case, the end result is usually a collection of fragmented sediment bodies in different passages of the system and it may be extremely difficult to reconstruct major portions of the history of sedimentation without detailed analysis of the greatest possible exposures in various parts of the cave. Bull and Carpenter, when discussing the predominantly water-laid deposits of Goatchurch Cavern and Sidcot Swallet (Mendip), stress the difficulties involved:

> The bane of any cave sediment study that attempts to reconstruct the past history of a cave is the lack of record. Often the very sedimentary phase that leads to the deposition of a sediment is responsible for the eventual modification or removal of that sediment. A fluvial cycle may both deposit and erode part, or the whole of a sedimentary layer, indeed the remnant sedimentary episodes left in a cave may represent a record punctuated more with gaps than sediments. Subsequent modification to the cave deposit may rework the material within the cave adding to it deposits, which themselves were reworked elsewhere in the system. The net result is the complete mixing of sediments that were both temporally distinct even within one small cavern, and spatially distant within and without the cave system as a whole. (1978:54)

The complexity of water-laid deposits in caves is well illustrated by the sophisticated techniques now thought necessary to interpret such deposits. A comparison of the procedures used by Ford (1964) at G.B. Cave (Mendip) and those used by Bull (1976, 1977a, 1978a) at Agen Allwedd (Powys) shows just how fast this discipline is developing. In the light of such research, it is

very difficult to accept the validity of studies of alluvial sequences, carried out on restricted exposures and interpreted in terms of changing exterior environment, such as that by Laville (1975) at the Grotte de l'Eglise (Dordogne).

The distribution of alluvial sediments within a cave tends to follow a general pattern which is useful when it comes to deciding how to handle the investigation of a given deposit. Influent caves, by definition, are graded to the local watertable. As a general rule, so are effluent caves, unless there is a good geological reason (usually associated with a radical change in bedrock lithology) why they should remain active when perched far above the normal surface drainage in the valley. This means that, with both influent and effluent caves, there is often a clear link with the local geomorphology that can be exploited even when very little is known about the interior of the cave system. As one moves deeper into the system, from either end, this link becomes increasingly obscur. The altimetric distribution of major passages and chambers may sometimes show a relationship with exterior geomorphology (cf. Sweeting 1950), but it is then necessary to relate extant sediments to periods of cavern development. As was noted above, when it is a question of interior alluvial deposits, only extensive knowledge of the system as a whole will permit a detailed understanding of individual exposures. It is therefore clear that it is most important to differentiate between near-surface deposits and interior deposits, since our expectations, at least with respect to the stratigraphic possibilities, will be different in the two situations.

Deep cave deposits often consist predominantly of alluvial sediments, the remainder being the result of speleothem

formation, mass movement and rockfall. In near-surface situations, alluvial deposits will often occur at the base of the sequence, with other types of sediment stratified above them. This would usually be interpreted as the normal downward evolution of the limestone drainage system. Reactivation may occur at a later stage. If reactivation is violent and short lived, a cave may be more or less totally flushed of older deposits, leaving only a jumbled remnant at the base of the following sedimentary fill, with no organised alluvial deposits at all. If reactivation is more sustained, flushing may still occur but a new alluvial sequence may later build up. Sometimes polycyclic sequences, composed mostly of alluvial sediments, may survive in any area of a cave, often due largely to cementation of older fills. A typical sequence might begin with coarse material passing upwards into well bedded sands and finally into laminated silts and clays, with stalagmite at the top and general impregnation by carbonates. Erosion at the start of a new cycle may cut a channel through the old fill but may leave better cemented deposits (often only the cap stalagmite) adhering to the walls. In wider passages, even uncemented deposits may survive, with a later misfit stream cutting a meandering channel through the old sequences and, sometimes, on into the bedrock below (cf. W.K. Jones 1971). With weaker reactivation, either in terms of duration or energy levels, more of any older non-alluvial deposits may survive, again especially if they are cemented. In this way, several alternations between alluvial and non-alluvial units may occur. It was stated above that the activity of influent and effluent caves may often be linked with exterior events. Thus, reactivation might be associated with the burial of previously active caves under glacial till, or with the formation of a lake.

However, such links must be demonstrated and not automatically assumed (cf. section 1.4.). In the absence of dating evidence, of clear links with climatic phases and/or demonstrable relationships with exterior morphostratigraphic units, it is very dangerous to interpret cyclicity of any type, when identified in a restricted exposure in any part of a cave, as anything more than variation in the local (passage) hydrology. Something as unpredictable as the fall of major roof blocks upstream could be responsible for the cessation, renewal and fluctuation of stream flow.

The relief of a cave area often has an effect upon the likelihood of a frequent repetition of major alluvial/non-alluvial cycles, or even upon more simple alluvial cycles. In areas of high relief, energy levels will be such that older deposits will more often be eroded by reactivation. Also, with high relief, the altitudinal difference between successive influent or effluent caves will tend to be greater, so that a more important 'event' (internal or external) is often necessary to cause reactivation, although the situation may be radically complicated by limestone lithology and structure and the degree of karst development. Complex interstratification of important alluvial and non-alluvial units is much more common in areas of low relief, especially where there is low precipitation (or general water availability). Such situations are rather rare in British near-surface cave sections containing deposits covering any great time span.

There are five main controls which govern alluvial sedimentation in caves:

(1) Frequent changes in passage width produce radical variation in flow speed independently of discharge rate.

(2) Irregular long profiles produce radical variation

in the gradient of vadose flow.

(3) The presence of flooded sections has a radical effect upon sedimentation.

(4) Varying discharge changes the flow rates and the distribution of flooded sections.

(5) Sediment in transport and pre-existing sediment interact with the passage morphology and flow regime.

Consider as an example a particular type of sump, with a section of passage where the roof is lower than the floor of the continuation of the passage on either side, a configuration usually due to a phreatic origin. The two branches of the sump may be near-vertical, only gently sloping or markedly assymetrical, arrangements often due to the structure of the bedrock (cf. Ford 1971). A vadose stream flows into this sump, so that the latter begins to fill with water. At first, the sump acts like a pool, producing a local base level which slows the flow. Sediment brought in by the stream is deposited, the coarse material first and then the finer material further into the sump. As the sump fills with water (overcoming any loss that might be occurring through cracks and small conduits communicating with the bottom of the sump), the level will eventually become high enough to cause an overflow on the downstream side. There will now be through-flow, with input equal to output (in a closed system). The cross-section of the sump will probably be greater than the cross-section of the actual stream channels on either side, so that flow will continue to be slower in the sump than in the stream proper. Sediment will still be deposited in the sump, but coarse material will penetrate further in and the finest suspension load will pass right through. As more sediment is deposited, the sump will become more restricted, so that flow

speed within the sump will rise in order to balance the input--output relationship. Thus, increasingly coarser material will pass through, and only the coarsest material will be deposited in the sump. The influent stream now goes into flood stage for some reason, as any stream inevitably will, and on a relatively short time scale. The resulting rapid flow will elutriate the existing deposits, washing out the fines and leaving only the coarse elements in the sump. Any clear coarsening upwards sequence, that we might have been able to interpret, is destroyed. In a stream, coarse particles are moved as bedload by sliding, rolling and saltation. However, coarse particles can only leave the sump if flow is actually strong enough to suspend them in the upward section. Usually, the coarsest material will remain in the sump until it has been abraded down to a size which the flow can transport. Now, if this cycle of normal flow, interrupted by flood stage, is repeated many times, the sump will accumulate a jumble of coarse material that never quite has enough time to become sufficiently comminuted for the flood stage to remove it. The sump will never be totally blocked and an equilibrium between particle size, remaining free cross-section and average flood velocity will be set up. Individual particles may leave the sump and new particles may enter, but there will be little if any coherent stratification. In some cases, coarse elements may become wedged together within the sump, so that even flooding cannot shift them and water flow is only possible through the interstices. Limestone debris, probably from local collapse events, is often involved in such blockages, referred to as 'ruckles' by cavers. Only when stream flow begins to wane in a more or less definitive manner (due to a new route or a reduction of water input on a long time scale) will fines be deposited in

the sump with any chance of remaining there. Fines will fill the major pore spaces between the coarse elements and will build up a stratified sequence above the coarse deposits, further restricting the available conduit. If weak stream flow continues for a time, the sump may become more restricted than the stream channel, so that initial silts may coarsen upwards into fine sands as flow speed in the sump increases, setting up a new equilibrium condition (cf. Renault 1968 for a discussion of equilibria in conduit flows). As flow wanes even further, the sequence may reverse to a fining upward trend. Eventually the whole sump may be choked with sediment, with water movement reduced to percolation. If this passage is later reactivated, water will build up behind the blockage until parts or all of the sediment suddenly fail under the increasing hydrostatic pressure. Some disturbed, dissected and elutriated remnants of older deposits, together with loose coarse elements, may remain in the sump, and the whole process starts again.

The presence of sumps in the majority of cave systems is the main factor responsible for the general fining inwards trend of alluvial sediments in caves (cf. Renault 1968). Many researchers have noticed the apparently 'rapid' (in terms of distance) decrease in angularity of stream gravels into caves. Certainly, the bedrock conduits and channels will increase the possibilities for abrasion, but it is the fact that gravel is trapped for long periods in sumps or other irregularities, coupled with the fact that water (carrying abrasive sand grains) must still pass over or between the gravel, that causes most of the rounding (cf. Newson 1971). This would appear to be supported by the findings of Bull (1978a) in Agen Allwedd (Powys), a large, low gradient system with many misfit channels and few sumps.

Here, rounding does not appear to be very significant, although apparent 'bulk' rounding increases downstream due to differential mixing of two sedimentary stocks. Despite the fact that many researchers have claimed that 'karstic' pebbles are flatter than those in surface streams (cf. p.408), the property of flatness would not appear to be particularly diagnostic. Indeed, in streams with irregular rock floors including small potholes, some pebbles may be almost perfectly spherical. The present author has never come across a situation where there was a significant difference between the shape of gravel in a torrential upland stream and in a cave fed by that stream, unless a change in sediment source occurred somewhere along the route.

In a restricted excavation of fossil deposits, it may not be particularly obvious that one is approaching a sump. On the upstream side, pond deposits may be quite extensive if the passage gradient is low. Nearer the sump, coarser clast-supported material may be present as gravel lags. If the sump itself is blocked by wedged coarse material (supra), a sieve deposit may develop on its upstream side. Particles of a given size are wedged in the sump, leaving voids of a smaller size. Through time, particles that are about the same size as the voids will work their way into the wedged deposit until they too are caught in slightly smaller voids. Average void size therefore decreases as smaller and smaller particles are caught in the 'sieve'. Unless the stream can remove the blockage, any material that is coarser than the current size of the voids in the sump deposit must be dumped upstream. A continuous lateral fabric trend is therefore produced. In the sump, most particles above sand size will be supported by particles that are as large or larger. Moving upstream, particles will be increasingly supported by other

particles that are increasingly smaller. There will be a sort of 'inverse fine-tailed grading' of the framework (but not necessarily of the overall particle size distribution), representing the increasing proportion of the total load that was forcibly deposited at the same time, whilst the voids of the 'sieve' became smaller. This simple pattern may, of course, become much more complicated if there are significant variations in flow stage and type of sediment input. On the downstream side of a sump, very little material may be deposited if the passage does not become significantly wider than the sump conduit itself. If, however, the passage widens out, the resulting drop in flow speed may allow the deposition of the coarsest material that is able to pass through the sump. A characteristic feature on the 'lip' of a sump, just as one enters the downstream passage, is a longitudinal gravel bar with a hump-back at the crest and signs that normal flow passed on either side, between the bar and the cave walls. Such bars may also develop under truly phreatic conditions (cf. Bögli 1980).

In wider passages, between restricting features such as sumps or narrows, more familiar alluvial patterns may occur. W.K. Jones (1971) recognises three main styles. Braided streams, usually occupying all or most of the width of a passage, occur at relatively low gradients, when a supply of non-cohesive sediment (mostly sand and coarser particles) is "over-abundant". Water flows over a thin layer of sand, with gravels in the main channels. In the examples quoted by Jones, braiding is localised, lasting for only about 30m. A meandering style develops at low to moderate gradients, where sediment input is less abundant. A stable pattern, with the formation of point bars, is only possible when the sediments of the passage are relatively

cohesive, giving firm banks. It would appear that well developed meandering is dependent upon a pre-existing sedimentary fill, although meanders often evolve further by cutting into the cave wall, and the whole stream may even erode a bedrock trench - a special form of superimposed drainage. The last style involves relatively straight V-shaped canyons, cut down into less cohesive sediments (usually silts and sands). Although such sediments are easily eroded, they may still form relatively stable banks that may be many metres high. The stream runs, at least initially, at a relatively high gradient but total discharge must be low to leave the banks intact.

It is not clear how much general validity Jones is claiming for the above observations, since the examples he gives are from restricted study areas. In a discussion of channel pattern in fluvial environments in general, Schumm (1977) recognises the following end members of a continuum of patterns. The first concerns streams carrying mostly suspended load. In a stable condition, flow cross-section is relatively narrow, sinuosity is high and gradients are relatively gentle. When there is surplus load, most deposition occurs on banks (channel narrowing) with only minor streambed deposition. When there is a load deficit, streambed erosion is dominant (entrenchment), with little initial channel widening. The other end member concerns streams carrying a high percentage of bedload. In a stable condition, flow cross-section is relatively wide, sinuosity is low and gradients are relatively steep. When there is surplus load, streambed deposition is dominant, with island formation. When there is a load deficit, channel widening is dominant, with little streambed erosion. These two descriptions do not correspond

strictly to meandering and braided patterns, but there is nevertheless a common association between, on the one hand, suspended load streams and meandering and, on the other, bedload streams and braiding. The types described by Schumm correspond in many details to the first two types described by Jones from caves. The main difference is in the gradients, especially with respect to the low gradient, braided patterns in caves. It is therefore necessary to insist upon the importance of high sediment availability and, most probably, of high (even forced) and irregular discharge rate in connection with braiding. Meandering should be associated with more regular, and possibly lower, discharge. The present author has often observed braiding in European cave streams, associated with coarser sediments, but very often the sediment cover is quite thin and there are patches of bare bedrock. This suggests that gradient is more often controlled by the cave passage and not more directly by the stream hydrology, as would normally be the case in the open. In an environment where only finer sediments were present but were available in large quantities, it is possible that high discharge events might set up a temporary braided pattern. Certainly, the stream in G.B. Cave (Mendip), which is in the process of removing recent debris flow deposits (cf. p.383), has adopted a braided pattern in silts and sands just before it reaches a choked sump. The third of Jones's types, linear V-shaped canyons, are dominantly erosive forms. It would seem to the present author that these canyons must develop very quickly, possibly by nick-point migration after a rapid drop in base level or after input of water to a previously dry section. Slower down-cutting should result in at least some depositional features perched on the canyon sides, as well as a degree of sinuosity.

The depositional structures developed in modern cave streams and pools have rarely been described in detail. In situations where channels (of any type), lenticular bedding and at least some coarser sediments are present - a combination representing quite an active environment - useful analogues might be found in studies of upland open-air streams that have some rock-cut channel sections, irregular long profiles and generally high average discharge (cf. Carling & Reader 1982). A very common form in caves is persistent laminated or planar bedding, involving sands, silts and/or colloids, with occasional coarser stringers. Such bedding may occur in stream channels (especially in wall-to-wall streams in narrower passages), in phreatic conduits and in pools of various sizes. Generally, this type of bedding represents comparatively gentle depositional conditions, but the exact environment can only be ascertained by extensive and detailed study. Laminations may often be graded, and may even show couplets resembling varves. The author has never come across a really convincing claim of varves (sensu stricto, i.e. annual rhythmites) in a cave; although the descriptive terminology may be useful (cf. Reineck & Singh 1980:123-130), the periodicity of any rhythmites is a characteristic to be proven, not assumed, and proof will nearly always be extremely hard to come by.

A good example of the study of rhythmites in a cave is provided by Bull (1981), who describes 30-40 mm of Late Devensian "cap mud", comprising 196 major laminae, from Agen Allwedd (Powys). The particles are all below 63 microns (silts and colloids) and the laminae appear to have been deposited by parallel accretion (settling through relatively still water, with fixation to existing sediment surfaces by inter-particle attraction stronger

than gravity). This is supported by the fact that 'micro-slopes' may exceed 70° with constant thickness of laminae and no signs of slumping, although grain size may be slightly coarser on steeper slopes. Laminae are remarkably persistent and there are no coherent lateral trends in thickness of laminae or in particle size along a kilometre of passage, suggesting the absence of any significant current. Under cracks and fissures in the rock, small cones have developed, with thicker and coarser laminae. Bull interprets these deposits as being due to multiple input (cf. the cones), with pulses of sediment being injected into essentially standing water; the water probably filled the passages completely, at least in some areas, since cap muds sometimes reach the roof. Bull concludes that these pulses were probably "microclimatically" controlled, but he avoids the term 'varve'. In an earlier publication (Bull 1977b), he had suggested that individual laminations, or pairs of laminations, at this site were certainly not varves, since the characteristic coarse-fine couplets of true varves were absent. However, he did raise the possibility that groups of c.30-60 laminations, showing roughly rhythmic variations in mean grain size, might represent annual cycles. It would appear to the present author that only by considering such factors as pollen or detailed chemistry could this possibility be pursued any further and, even then, with no guarantee of a decisive result.

The depth-of-water problem is often difficult to solve in caves, especially when only restricted exposures are available. The maximum possible depth in a given passage is obviously the distance between the sediment and the passage roof, but the total hydraulic head may be much greater. A passage may be totally blocked by alluvial sediments, proving that water reached

the roof at least in the later stages. Some erosional features cut into the rock of the roof (inverted roof channels, roof scallops, anastomoses, dome-pits, etc. ; cf. Warwick 1976; Bögli 1980) testify to complete flooding at some time in the past, but it will be difficult to relate such features to extant sediments without extensive study of the cave system. In some cases, very fine sediments may build up by parallel accretion around the entire periphery of a flooded conduit, but such material will not survive dewatering for very long unless the central space is first filled with normal sediments. If the level of a free water surface remains constant for a significant period (a situation most commonly arising in pools), a number of traces may survive. Walls may show quite extensive horizontal notches (cf. p.793), corresponding to the water surface; these are not, of course, present on rock covered by older sediments. The water may control the growth of stalagmite, with wall curtains stopping at the water-line but swinging out to give various horizontal forms (cf. Chapter 13). The roughly horizontal tops of some micro-deltas, formed where feeder flows enter a pool, occur just below the water surface. Bögli (1980) has suggested that, whereas decomposed organic matter ('humus') has a greater density than water and will therefore sink, structured vegetable material (seeds, leaves, woody debris, charred fragments, etc.) will float and will therefore be concentrated along banks close to the water-line. In deep pools (5m and more), it may be possible to recognise a gradual drop in total organics with depth. Except in the case of charred material, the interpretation of organic matter in this way will depend upon the absence of oxidising conditions at all times between initial deposition and final excavation, or at least upon arguments that suggest that all

levels of the deposits have suffered a similar, and not too extreme, degree of oxidation.

Fluctuations in water level, with partial or complete drying of some sediments, may leave traces. Cracking of fine sediment does not always indicate absence of a water body (cf. synaeresis, described on p.274) but, if curved platelets ('mud scales') are also present and cracks are relatively deep and linear (as opposed to curvilinear), emergence and shrinkage is indicated (cf. Bögli 1980). Partial drying may induce deposition of iron hydroxides and carbonates which, in a fluctuating system, may result in quite extreme colour banding. Fluctuating water content within a sediment, especially when water input is quite rapid, may produce vesicular structures as air is expelled. This may occur at a period much later than the original deposition of the sediment (cf. the discussion of "birdseye structures" by Bull 1975). Surge marks are features superficially resembling rilling but which, in cross-slope vertical section, are seen to be composed of parallel laminae with an undulating or zig-zag pattern, and are thus not dominantly erosional features. Bull (1978c) discusses these structures in some detail and notes that they often form upstream of constrictions or, in one modern case, on a mudbank near a tidal sump affected by frequent but low amplitude surging. Although these features must often form at or about mean water level, Bögli (1980) claims that they may also be present in phreatic situations, so that some independent evidence for a free water surface should be sought.

Some indication of hydrological variables can be gained from the study of the textural parameters of alluvial deposits, as long as no spurious generalisations beyond a given exposure are claimed. Mean grain size can be related to mean flow speed

via Hjulström (1939) and Sundborg (1967) diagrams (reproduced in many texbooks, e.g. Reineck & Singh 1980). The sorting of a deposit also has a relationship with the regularity of flow. Care should be taken in the interpretation of both the finest and the coarsest sediments. Rising alkalinity will cause clays and some fine silts to flocculate and therefore behave as larger particles; flow velocity of a stream depositing very fine material may thus be considerably higher than theoretical consideration of individual particle sizes would suggest. Flocculation is well known from estuarine environments, where flocs may sometimes become stabilised to give a lasting structured sediment (cf. Reineck & Singh 1980), but there is no available information to suggest whether or not this might be possible in caves. During rapid erosion of clayey sediments, material may be transported as mud clasts of sand or larger size. Close inspection of an apparently massive deposit may reveal such clasts (cf. p.1146), which are originally spherical but which are usually compressed after deposition to oblate or even lenticular forms. Gravel beds may also be misleading with respect to contemporary flow speeds since they may represent lag deposits, derived from much older sediments, that have been formed by relatively weak flow that was only able to remove fine particles. Some idea of the competence of the 'youngest' flows can be gained by observation of the degree to which the gravel is now structured (e.g. imbrication, bars, etc.). In sediment-cut channels, bank deposits should be checked for an immediate source for channel gravels. Arguments concerning the link between texture and flow speed may also be useful in assessing the cause of erosion surfaces. For instance, Bögli (1980) describes a situation in the Hölloch Cave (Switzerland) where gravels overlie an erosion

surface cut into clayey deposits; since the flow velocity that would be necessary to erode the clays is similar to the velocity which would have allowed deposition of the gravels, Bögli concludes that erosion and gravel deposition were syngenetic. As long as an approximation of water depth could be calculated, a consideration of bed forms might furnish even more precise information on flow regimes (cf. Simons & Richardson 1962; Allen 1968; Reineck & Singh 1980). White and White (1968, 1970) give further details concerning hydrology and alluvial sediments in caves.

One last situation, which is quite common in caves, should be mentioned, namely the occurrence of major sequences of relatively poorly structured, gravelly sediments, comparatively close to the point of entry into the cave system (e.g. the lowest deposits at Pontnewydd Cave, p.1190). The difficulty here lies in differentiating between 'torrential' fluvial deposits and various types of mass movement (cf. Chapter 9). Fast mass movement is often associated with water flow, so that occasional current-induced structures do not prove a fluvial origin. Conversely, generally poor sorting or high C/M values (the ratio between the grain size of the coarsest percentile and the median grain size) sometimes occur in fluvial deposits due to periodic high discharge. Also, little is known about coarse, subaqueous mass movement, which might be expected to show 'intermediate' features. If a badly sorted deposit with relatively homogeneous fabric also shows extensive matrix support for coarse particles, it almost certainly represents mass movement. However, with decreasing fines and more clast support, it becomes more difficult to rule out the possibility that the fines were later introduced into a 'clean' stream gravel. Sometimes, a pair of clasts may

show mutual erosive 'cups', formed as loosely trapped particles rub against each other due to vibrations set up by a current of water flowing through the gravel. Even in this case, the deposit might still have been introduced by mass movement, reduced to a gravel lag by water flow (with 'cup' formation) and finally infiltrated with fine sediment. Only in the case where a sieve structure, similar to that described above (p.428) but with a vertical rather than a lateral trend, is apparent can infiltration be proven. The situation in caves is closely analogous to some exterior alluvial fans, where every gradation between fluvial and mass movement deposits may occur (cf. W.B. Bull 1977). Often the fact that fast mass movement and torrential flooding imply similarly disruptive environments may mean that the difference between the two is of little contextual importance. However, if possibly discrete geological, biological or archaeological assemblages are present in individual strata, or if detailed correlation with other deposits inside or outside the cave is required, the difference may become more important. In this case, very detailed work will be necessary and more extensive exposures must be studied.

In conclusion, it must be reiterated that alluvial deposits in caves may provide useful information concerning the localised sedimentary contexts resulting in a given exposure, but that extensive generalisation beyond that exposure nearly always involves widening the scope of research, in terms of both detail and study material, by several orders of magnitude. In a restricted excavation, of the type most commonly undertaken at an archaeological or palaeontological site, the climatic, and even the geomorphological, implications of alluvial deposits will often remain obscure.

10.3. Sub-surface Processes

The processes considered briefly in this section result in the translocation of particles within a sediment body. All such processes are, strictly speaking, post-depositional, but the time scales involved may range from almost zero, with significant particle movement occurring immediately after initial deposition, to 'geological', with particle movement best considered as a form of diagenesis (cf. Chapter 16).

Consider a lattice consisting of identical large spheres, a lattice of infinite two-dimensional extent but only one sphere thick. If groups of four spheres were placed so that the centres of the spheres fell on the corners of a square (so-called 'cubic' packing), the ratio between the diameter of another set of mobile spheres (the largest spheres that could pass through the lattice of original spheres) and the diameter of lattice spheres would be 1:2.41. If, however, the lattice was reorganised so that groups of three spheres had centres falling on the apices of an equilateral triangle (so-called 'rhombic' packing), the diameter ratio of mobile to lattice spheres would become 1:6.46. As the packing of a sediment changes, so does the size of the voids between particles. In the present context, it is the 'maximum connecting void network diameter' which is of interest, not general porosity. If, over the lattice described above as having 'cubic' packing, were placed an identical lattice in such a way that groups of eight spheres had sphere centres falling on the corners of a cube, the ratio of interest here would still be 1:2.41; although the void centrally placed with respect to the eight lattice spheres would allow a ratio of 1:1.37, the largest particle that could be placed in this void would not be mobile.

The simplest case of particle movement arising in real sediments may be called 'through-transport'. A fine sediment, originating elsewhere, may pass right through a clast-supported coarse bed. The controls of this process would simply be the ratio between the diameters of the fine particles and those of the interconnecting voids (a ratio which must by definition be lower than 1:1), together with the nature of the motive power. Again working with theoretical spheres, Statham (1974) has calculated that, for sediment to move by gravity alone in a dry context, the diameter ratio of mobile spheres to lattice spheres needs to be as low as 1:15. With natural particles, even allowing two populations each containing particles of nearly standard size, both overall shape and surface roughness will necessitate an even lower ratio. As water is added to the system, it will act both as a transporting agent and a lubricant. At higher flow velocities, the only significant limiting factor will be void size, although, at very high velocities, vortex formation (a 'plughole' effect) may begin to inhibit movement.

Through-transport may be reduced in two basic ways. First, a natural clast-supported bed will not have an absolutely regular network of voids. If the finer sediment which is being introduced has some relatively coarse particles that can get a little way into the lattice but then become trapped, overall void size will decrease. This clogging process may result in the sieve deposits discussed in the last section (p.428). Soil engineers create such a situation artificially, in order to protect the boundaries of sediment dams and banks; the lattice material in this case is called a 'filter'. Bowles (1979) cites certain criteria that must be met in order to produce the best filter. The filter material must be poorly sorted, which ensures

dense packing and thus a fine and stable initial void network. Three empirical criteria relate aspects of the particle size distributions of the filter (F) and of the 'soil'(S) which must not pass through the filter. The notation, DF_{50} and DS_{50}, means the diameter of particles below which 50% by weight of the size distribution lies, for the filter and the 'soil' respectively. The criteria are:

$$(1)\ \frac{DF_{15}}{DS_{85}} < 4 \text{ to } 5;\quad (2)\ \frac{DF_{15}}{DS_{15}} > 4 \text{ to } 5;\quad (3)\ \frac{DF_{50}}{DS_{50}} \leq 25$$

The first and third criteria suggest that the filter should not be too coarse in comparison with the 'soil', whilst the second suggests that the finer tail of the filter distribution should be coarser than that of the 'soil' distribution. This second criterion ensures that, once the filter is in place, it will become firmly welded to the 'soil' by infiltration of fines. Natural deposits of this type are not restricted to the true sieve deposits of alluvial environments. The tops of many scree bodies in caves show this effect to some degree. For a given pair of upper (or up-stream) sediment and lower (or down-stream) sediment, the regularity of development of the clogged or welded zone is a function of transport energy (from gravity, to percolation, to stronger flow).

The second way in which through-transport may be reduced is by interference from a bed underlying the open-lattice bed. If this lower bed has a restricted void network which will not allow the passage of significant sediment, material will simply build up within the upper bed. This is a very common occurrence with cave screes underlain by finer beds. Again, the transport energy will control the final result. At low energy, fill in

the framework voids will be very loose with common airspaces (especially if the fill material is badly sorted), whilst at higher energy, the fill will be denser and more continuous (especially if the fill material is well sorted and relatively fine). Note that, even with a truly clast-supported framework, it is not necessarily obvious whether the fill totally post-dates the framework or whether it is roughly, or in part, syndepositional. This is particularly true in cases where the fill is of medium grain size and is intrinsically mobile (non-cohesive coarser silts and sand). It is therefore important to look for large filled 'voids' in the framework that contain a large particle, which is too large to have arrived through neighbouring voids and which is supported by fine matrix and not the framework elements. Even in this case, care must be taken not to misidentify particles which have simply broken off nearby framework elements. The use of allochthonous rocks and animal bones avoids any uncertainty. There is no way of proving that none of the finer material in a clast-supported framework is syndepositional, unless the framework and fill certainly represent radically different energy environments (e.g. a well structured fluvial gravel filled with very badly sorted sediment including small, angular limestone fragments). Theoretically, it is possible for the fill and the framework to be totally syndepositional (e.g. on a beach, with the surface of the framework pebbles always building up just above the accreting surface of fill sand which is thus protected from the wind) but, in most cases, it is safer to assume some infiltration of fill at a later date.

So far, we have considered situations in which the mobile material is derived from 'beyond' the framework bed. Sometimes the finer sediments represent 'initial' deposition at the end of

a major transport process, as in the case of wash infiltrating a scree body and there depositing its load. Often, however, the fines are derived through sub-surface translocation of pre-existing sediment, either from an overlying unit or from an original matrix within the unit under study. The processes of sub-surface erosion must therefore be considered.

The process of elutriation is the separation of a mixture into components according to the energy level of a passing fluid (air or water). The term is best reserved for relatively high energy situations in which a limiting property of the material which remains, the 'lag', is clearly related to the fluid process. Thus, a clean gravel lag in a stream bed will have had its particle size distribution truncated, more or less abruptly, at a size corresponding to the maximum erosive power of the elutriating current. Even if the lag is refilled with fines at a later date, there may still be a tell-tale gap in the overall size distribution. Elutriation may occur to a significant depth in stream deposits, but the current will eventually lose power and coherence so that this process is usually a near-surface effect. Elutriation will usually disturb the coarser elements, at least slightly, and it may often involve considerable settling. The process of suffosion is the removal of particles from a sediment by through-flow of water (the correct spelling is 'suffossion', but the other form is now firmly entrenched in the literature). The word means 'undermining' but geological usage simply implies removal of fines, not necessarily with any significant subsidence. Suffosion is usually a relatively gradual and prolonged process, operating under fluctuating energy conditions which it is not normally practical to reconstruct in any detail. As long as particulate sediment is present and pore

water moves, there is no depth limit for this process. The term 'lag' is sometimes used in the context of suffosion; it applies, not to the whole coarse framework, but to extremely localised pockets of re-organised coarser matrix material (silt, sand or the finest gravel) that are very roughly related to the threshold competence of through-flow. If little material has actually been removed from a deposit even though fines have demonstrably moved within the deposit, the general term 'translocation' is best applied, with any necessary qualification. It may be possible to define zones of depletion and zones of concentration, but the cognate terms 'eluvial horizon' and 'illuvial horizon' are best restricted to true pedological phenomena (cf. section 12.2.).

Before considering the structures produced by various types of sub-surface erosion, it is important to recognise the separate controls involved in mobilisation and transport. In an unrestricted situation, once mobilisation of particles has been achieved, significant transport is usually guaranteed, since the energy required for transport is less than that required for the initial erosion. Below the surface, transport can only occur if there are favourable relationships between the mobile sediment and the coarse framework, as discussed above. Conversely, the framework may be conducive to movement, but high matrix cohesion may inhibit mobilisation. Cohesion is largely a function of colloid content and cohesive sediments will tend to be less generally permeable. Water flow may therefore be concentrated, in cohesive matrices, along any discontinuities (secondary permeability), thus raising the localised energy level and partially compensating for the resistance to erosion.

Let us consider first the water input end of the system. A simple example would be a persistent drip falling on a roughly

horizontal sediment surface. There would be various superficial effects, such as a crater or a slightly coarser 'splash lag', associated with the impact energy. Below the surface, flow energy would decline rapidly. If the fine matrix was very well sorted and reasonably well compacted, there would be practically no sub-surface movement of particles, since there would be no available voids of the requisite size, no matter how coarse the clast framework might be. If the matrix was badly sorted but still quite compacted, minor translocation might occur but clogging would very quickly limit the effect. We may therefore conclude that significant sub-surface erosion is not initiated at the point of water input. The author has observed many modern drip-induced features in compacted sediment and the sub-surface effects are usually extremely limited (under 10cm, and often as little as 1cm), with slightly more penetration under more active drips. Colloids and carbonates, introduced by the drip itself, may penetrate a little more but they also help to reduce further any movement in the original matrix.

Sub-surface erosion starts at points where water leaves a sediment body or where water passes from a matrix-clogged context to one with abundant void space. Erosion then proceeds back up the flow path of the water, removing the matrix and making available the full potential of the coarse framework. It is not necessary that the sediment under attack have an autonomous coarse framework in the first place. Matrix-supported sediment may be affected as long as the voids, resulting when larger elements settle after removal of the matrix at the erosion front, are sufficiently large to allow the passage of more matrix derived from further up or back in the deposit. The most important prerequisite is therefore marked bimodality in the particle size

distribution and not the dominance of coarse elements, although there must be enough coarse elements to set up a significant void network relatively quickly. Suffosion of a matrix-supported deposit obviously involves subsidence, an effect that can reach impressive proportions over long time scales (cf. Chapter 16).

Suffosion of a deposit may sometimes be widespread, affecting an extensive tabular body, but the author has never seen this in a cave deposit except where water is escaping at a nearby free surface. An enormous volume (probably in excess of 2,000m^3) of the deposits at Bacon Hole (Gower) has suffered partial, and in places total, removal of fines, because the sequence is cut to bedrock by a 10m high erosional cliff produced by undermining of the outer deposits during the holocene rise in sea level. Normally, suffosion is rather patchy, with irregular but roughly linear openwork features. Markedly linear development is often referred to as 'piping', although, strictly speaking, a pipe is a cavity that is considerably larger than the normal voids of the framework. True pipes are best developed in cohesive and well structured finer sediments (especially in seasonally arid environments and in steeply sloping terrain; cf. Bunting 1961; A. Jones 1971). The author has never seen this form of erosion in fine sediments in a cave, but true pipes may develop in coarser material if stability is raised by point cementation of the surrounding framework clasts. At Bacon Hole, pipes originating at the erosional cliff are penetrable for up to 2m and then continue out of sight as more restricted cavities, with 'stalactites' on the roof. Such true piping is obviously dependent upon some large clasts being able to move 'downstream', a situation that will arise in practice only when suffosion is acting from a completely free sediment face. An extreme example of piping,

resulting in a "tunnel" with a surrounding cemented zone, has been reported from the Qafzeh Cave (Israel; Bar Yosef & Vandermeersch 1981).

The development of suffosion may be more or less vertical in better drained deposits, with bulbous sections through less cohesive beds and restricted 'necks' through more cohesive beds. Such forms are still recognisable deeper within Bacon Hole, where they may perhaps pre-date the more general suffosion. In sloping deposits, with slope-bedded coarser elements, suffosion routes may pass diagonally through the deposit as a compromise between gravity and fabric. However, a suffosion route lies between the points of water input and output which do not necessarily have anything to do with the stratification. For instance, a cave containing predominantly down/out-sloping sediments may show a suffosion route counter to the stratification if wash arrives through the entrance but exits through a fissure under the sediments further back in the cave. Where percolating water reaches a more cohesive and impermeable bed, it may travel laterally before finding a route through that bed. Suffosion working back up the route will create a similarly 'stepped' feature. This explains the common occurrence in vertical sections of suffosion features with no apparent upward or downward links (cf. Sun Hole, p.907), features which must not be mistaken for 'stratified' phenomena. Suffosion may remain confined within a single stratum for considerable distances if the underlying sediment or bedrock has only very low permeability. If lateral variation in the overlying sediment is such that it too becomes impermeable, there is scope for the development of an artesian system, with upwelling of water, although the author has yet to see a convincing example that was not associated with a true

spring, involving a head of water developed in a whole bedrock conduit (cf. Rhinoceros Hole, p.988; also, the rupturing of stalagmitic floors, p.286).

At the base of more horizontal sections of suffosion features, discontinuous lags may develop which may sometimes comprise well bedded finer material (cf. Sun Hole, p.909). A more common structure involves small lenses with graded bedding, usually based on fine limestone clasts and resembling tiny sieve deposits with a vertical development. These structures are normally found near the lower boundary of the suffosion feature, but they sometimes occur 'perched' higher up in an open framework.

Suffosion, and indeed translocation of most types, is a disruptive process which imparts very little useful information. It is extremely difficult to 'date' the onset of erosion. Consider a situation in which an upper bed A overlies a lower bed B. A suffosion feature, originating deeper in the sequence, passes through bed B and penetrates a little way into bed A. Has the feature developed comparatively recently, with its upper limit representing a continuously rising erosion face? Or is it a fossil feature, older than the A/B boundary, which has been recently reactivated to penetrate a little way into bed A? The dating of speleothems within a suffosion feature may help, but this still only produces a terminus ante quem. The author has never seen a clear suffosion feature in a Middle or Upper Pleistocene cave deposit that has been completely refilled with fine sediment. This suggests that such preferential sub-surface drainage routes are part of the general diagenetic process and that they will not usually cease to function before the stage of lithification or total destruction. Nevertheless, there is no absolute reason why refilling should not occur and the possibility

should be borne in mind. Suffosion is a function of the spatial relationships between water input, water output and the intervening sediments; it has hardly any environmental significance (infra), even if the activity can be dated to a usefully short period. Suffosion must not be ignored, however, since it is a process which may cause the translocation and mixing of small bones and artefacts, pollen and mineral grains. The author has rarely seen a cave or shelter deposit in Europe that does not contain significant suffosed zones. It is rather surprising that reports in the literature are all but absent.

A type of translocation, called "thermo-suffosion" by Embleton and King (1975), operates under the same environmental controls as gelifluction (sensu stricto, cf. p.329) but where the sediment texture is more susceptible to suffosion than to mass movement. Typical thermo-suffosion requires permafrost and results in collapse features expressed as surface depressions. In the cave context, where normal suffosion is so common, absolute proof of permafrost would be necessary to support a claim for thermo-suffosion (cf. section 8.2.). Nevertheless, the concentrated water supply produced by melting of pore ice in cave sediments, and by general melting in the local environment, could well encourage suffosion, although it is difficult to see how this could be proven in ancient deposits. Near surface suffosion, in conjunction with frost heaving and snow overburden, may also give rise to 'stone pavements' (cf. Washburn 1979). These are closely fitting mosaics of horizontally bedded, large stones, with smaller stones in any interstices but no (pre-burial) fine matrix (removal by suffosion); the underlying sediment is poor in stones, since these have been sorted out to form the pavement (cf. section 8.2.). Permafrost

is not necessary but, since other agents (including man) can form such structures, proof of the involvement of ground ice is required.

10.4. Marine Processes

Coastal caves often contain deposits with marine associations which, because of the possibilities of widespread correlation, are of the greatest stratigraphic interest. The western coasts of Britain have many caves, but two of the best known areas are perhaps Gower (limestone basement), with a large number of palaeontological sites associated mostly with deposits laid down during and between the ipswichian high sea levels, and the Oban area (metamorphic basement), with various flandrian strandlines associated with occupations during the Late Mesolithic and Early Neolithic. However, it is more or less the entire periphery of the Mediterranean that boasts the most impressive array of coastal caves in the world, with archaeological sites spanning at least the last 750,000 years and palaeontological sites reaching back into the Pliocene.

Marine deposits, like fluvial ones, have been extensively studied and they will not be discussed in detail here (cf. King 1972; Reineck & Singh 1980). What is of interest is the nature of the marine association evidenced in a given deposit and the relationship between this and sea level. It is possible that a marine influence might be present in cave sediments that were deposited at altitudes of about 10m below sea level to about 50m above sea level (shallow marine, intertidal, storm beach, dunes, etc.). Since recognition of different transgressions even on the comparatively coarse scale of interglacials will usually require

resolution in the order of 5m or less, it is clear that we must identify the exact coastal sub-environment and that we must be able to relate this information to mean sea level.

The sedimentary record of changes in sea level close to coasts is never complete because the high energy levels of the breaker zone, coupled with longshore currents, ensure constant reworking and erosion of older deposits. One of the main reasons for the comparative richness of the cave deposits of the Mediterranean is that this is essentially an inland sea, with subdued energy levels stemming from low tidal amplitude and restricted fetch. The deposits of a transgression will have little chance to survive, whilst those of a regression are more likely to remain since they will be less often subject to immediate reworking.

On a lowland coast, with no resistant rocks, the various littoral environments will show good development and lateral differentiation, with a coastal zone perhaps as much as a kilometre wide. As sea level fluctuates through time, the position of the shoreline may change quite radically and the zone of alternation between marine and terrestrial sediments may be tens or even hundreds of kilometres wide. Other things being equal, such lowland coasts will accumulate the most coherent sedimentary sequences, not only because the facies are well differentiated in the first place but also since the erosive power of the sea will be concentrated at given geographical points for only comparatively short periods.

In contrast, most caves will be found in rocky and cliffed coastlines. Cliffs represent the current equilibrium between the erosive power of the sea and the resistance of the rock. The retreat of cliffs in rocks such as limestone is often

extremely slow and may average only a few metres or tens of metres over a hundred thousand years. The cliff limits the landward extent of successive transgressions, causing a more or less extreme reduction in the width of the zone of marine/terrestrial alternation. The sea reaches the cliff comparatively early in the transgression and remains there until comparatively late in the regression. There is much less scope for the lateral differentiation of facies and those deposits which do form are vulnerable to marine erosion over long periods. Coastal caves will therefore tend to contain very patchy sediments, usually laid down at or after the time (but not necessarily at the altitude) of the highest sea levels. Continuous temporal sequences may be very short and deposits of radically different age may be intimately associated in space. This effect is accentuated because of the 'historical accident' of our own position as observers during a period of relatively high sea level. With rare exceptions (cf. Spalding & Mathews 1972), the zone of marine/terrestrial alternation, already reduced by the cliff, is only easily accessible to us over a few tens of metres at its landward margin, and thus only in its 'warm' climate expression. This general model is, of course, complicated by the fact that successive transgressions do not occur to the same altitude. A series of transgressions, each weaker than the last, may leave a more coherent sedimentary record, whilst a single major transgression may destroy all pre-existing deposits. If the area in question is tectonically unstable, the situation will become even more complex.

Modern sediments, usually of relatively minor extent, that are being deposited in the intertidal zone are easily recognised in many low lying coastal caves. Various combinations of gravel bars, shingle spreads, pool deposits and laminar or

cross-bedded sand units are common, as are signs of biological activity. However, traces of such deposits in ancient contexts are rather rare. A good example from Britain is provided by the Inner Beach at Minchin Hole (Gower), which comprises well stratified and compacted coarse sands with shingle lenses and a middle shore molluscan assemblage (cf. Bowen 1977b, Plate 1, p.229). To the present author's mind, the age of this material is uncertain, although it must be at least as old as the main ipswichian transgression (cf. p.1095). Examples of various ages are more common from the Mediterranean, although structural details are not often described (cf. Miskovsky 1974; Le Vallonnet, Le Lazaret, Madonna dell'Arma, La Grotte du Prince, etc.). The rarity of fossil intertidal deposits in caves is not surprising since the encasing rock prevents dissipation of storm energy and promotes extreme turbulence. Many coastal caves show signs of marine action (planation surfaces, wave-cut notches, borings of marine organisms, etc.) without any accompanying marine sediments or, at best, with only a few disturbed pockets deep within fissures in the rock (cf. Blanc 1921). Pre-existing deposits may also show erosional benches, borings or even 'cup marks' produced by vibration of larger elements (cf. Hills 1970). Some caves appear to have been formed or at least heavily modified by marine erosion (e.g. the Kebara and Tabun Caves, Mount Carmel; cf. Horowitz 1979), but it seems likely that the designation 'sea cave', with its connotation of marine origin, has often been applied rather uncritically. The most extreme example of a marine association is furnished by the sequence at the Cunette Site at Sidi Abderrahman (Morocco), which actually includes littoral sediments deposited in 'caves' (deep marine notches and pseudo-karst) cut into older, cemented littoral sediments

(cf. Biberson 1970).

The 'marine' deposits most frequently recorded from cave sites are usually referred to as 'beaches'. The only common denominator for this class of deposits seems to be the presence of shingle, pebbles and/or cobbles, together with some marine mollusca. That such deposits are nearly always indicative of the proximity of the sea is beyond question. However, their altitudinal relationship with contemporary sea level is not always obvious. Modern shingle spreads may be seen on the Gower coast even up to c.20m above mean sea level. The presence of only a few pebbles or larger shells is even more equivocal since, not only can a storm sea breaking against a rocky coast throw such objects considerable distances, but birds such as gulls also collect them near nesting sites. Marine erosion of ancient scree-rich sediments may produce deposits that superficially resemble beaches. If an ancient sediment is well consolidated or cemented, the sea will cut an oblique or vertical erosion face. Fines may be removed from the zone just behind the erosion face and limestone clasts in this zone may be weathered to rounded or even truly spherical shapes. Such deposits are commonly found at the base of stratified sequences, resting directly upon the rock platform (cf. p.1048); demonstration of their true nature requires excavation beyond the zone of alteration. When quantities of true beach sediment are clearly present, it is still necessary to demonstrate an <u>in situ</u> marine constructional feature. Beaches are not haphazard piles of sand and pebbles. Many erosional agents, marine or not, may move beach gravel downslope at any point in a stratigraphic sequence, although beach gravel lags are particularly common immediately above bedrock. Such deposits only record a minimum altitude for an earlier beach, or beaches,

an altitude which may be higher or lower than the original sea level itself.

Wind-blown sand deposits, containing comminuted shell, foraminifera and other marine fossils, are commonly associated with coastal sequences. Such deposits are most likely to date from a regressive phase, since freshly exposed marine sediments are then available as a source (cf. Fairbridge 1971). However, it is dangerous to make a simple equation between a sand body in a cave and a regression if there is no supporting evidence. At least the top few metres of dune sands at Pennard Burrows (Gower) are due to storm events during the recent past, including the major event of 1528 in which the local abbey was completely buried. These dunes reach up to over 50m O.D. and penetrate over a kilometre inland. Even if it could be convincingly argued that we are at present in a significantly regressive phase, and that no such phenomenon could possibly arise during a transgressive phase, the dunes would still represent disjunctive sedimentary events and not a marine regression as such. Sand bodies resulting from the approach of the sea during a transgression are in fact known, such as those at East Guanagat Cave (Butzer & Helgren 1972) and Die Kelders I (Tankard & Schweitzer 1976), both in Cape Province, South Africa.

The interpretation of coastal cave deposits in terms of sea level is far from easy (cf. Chapter 23). In the past, researchers appear to have been too eager to interpret this or that 'raised beach' in terms of this or that 'interglacial'. Not enough weight has been given to the fact that coherent littoral sequences are demonstrably rare, indicating that most of those deposits which have survived must represent only a tiny proportion of the time involved in any marine hemi-cycle. Given

the further banal fact that the sea is an agent that can, and often does, create or destroy sediment bodies of tens or even hundreds of cubic metres literally overnight, it can be seen how incomplete and arbitrary the record is likely to be. Two 'beaches' separated by terrestrial deposits may indicate an interval of a hundred years or of a hundred thousand years, and the higher 'beach' may represent a lower sea level because of factors such as greater storminess, more direct prevailing winds or more conducive rock platform micromorphology. Accurate sea level data is unlikely to be available from a single cave sequence and a regional study, designed to smooth out the effects of exceptional events and localised conditions, will be necessary. Obviously, in order to avoid circular arguments, those published correlation schemes that rely heavily upon altitudinal data should not be used uncritically.