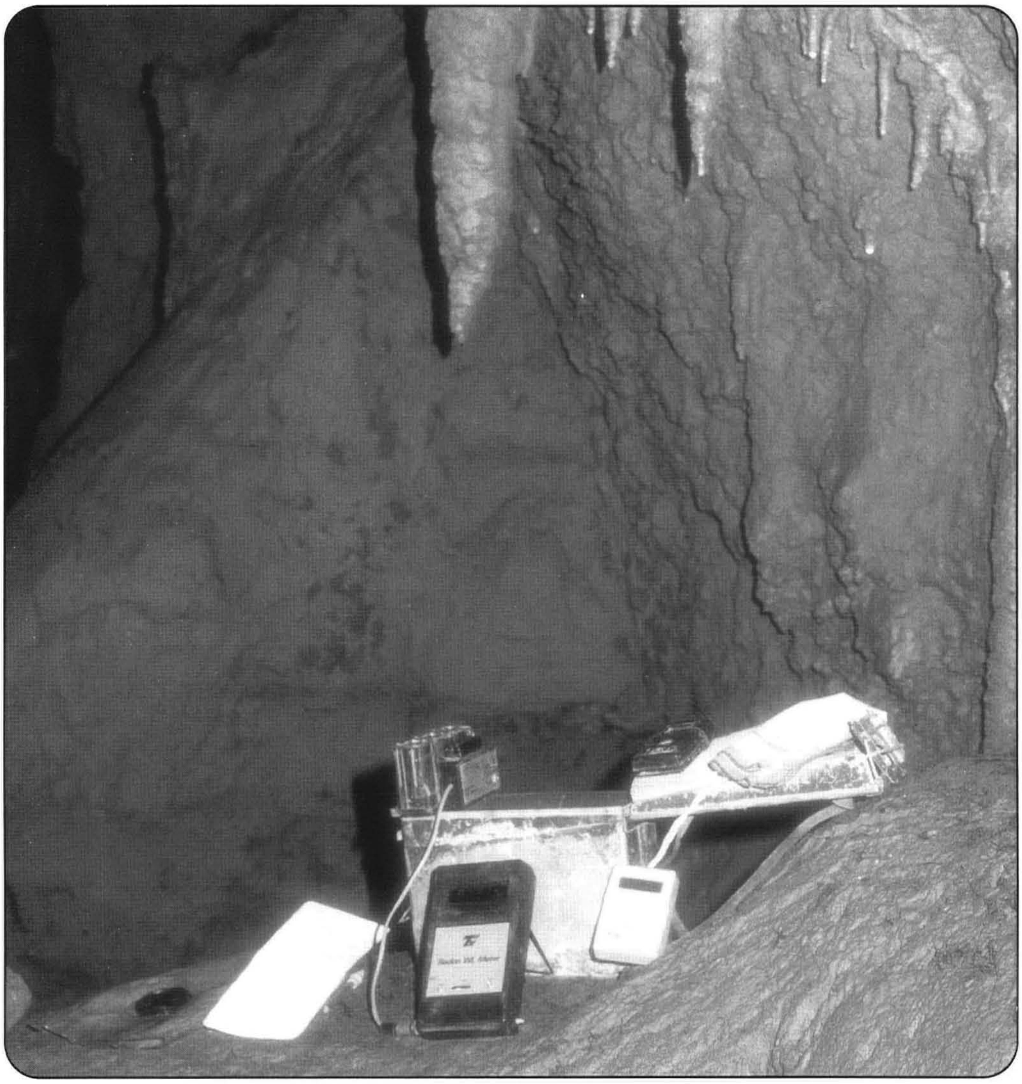
Cave and Karst Science

The Transactions of the British Cave Research Association

BCRA

Volume 23 Number 2 October 1996



Radon exposure during underground trips

Tafoni in the Maltese Islands

Karst and pseudokarst

3-D Vector Processing

Radio-location

Forum

Cave and Karst Science

Authors are encouraged to submit articles for publication in the Transactions of the British Cave Research Association under four broad headings:

1. Mainstream Articles
Scientific papers, normally up to 6,000 words, on any aspect of karst/speleological science, including archaeology, biology, chemistry, conservation, geology, geomorphology, history, hydrology and physics. Papers should be of a high standard and will be subject to peer review by two referees.

2. Development Articles
Shorter papers, normally 500-3,000 words, on aspects of karst/speleological science listed above, or more descriptive material such as caving expedition reports and technical articles. These will be reviewed by the editorial board unless the subject matter is outside their fields of expertise, in which case appropriate expert assessment will be sought.

3. Forum
Personal statements of up to 1,000 words on topical issues; discussion of published papers and book reviews. Statements should put forward an argument and make a case, backed-up by examples used as evidence.

4. Abstracts
Authors (or supervisors) of undergraduate or postgraduate dissertations on cave/karst themes are asked to submit abstracts for publication. Please indicate whether the thesis is available on inter-library loan. Abstracts of papers presented at BCRA and related conferences or symposia will also be published.

Manuscripts may be sent to either of the Editors: Dr. D J Lowe, British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK, and Professor J Gunn, Limestone Research Group, Department of Geographical and Environmental Sciences, The University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK. Intending authors are welcome to contact the Editors, who will be pleased to advise on manuscript preparation.

Notes for Contributors

These notes are intended to help the authors to prepare their material in the most advantageous way so as to expedite publication and to reduce both their own and editorial labour. It saves a lot of time if the rules below are followed.

All material should be presented in a format as close as possible to that of *Cave and Karst Science* since 1994. Text should be typed double-spaced on one side of the paper only. Subheadings within an article should follow the system used in *Cave and Karst Science*; a system of primary, secondary and if necessary, tertiary subheadings should be clearly indicated.

Abstract: All material should be accompanied by an abstract stating the essential results of the investigation for use by abstracting, library and other services. The abstract may also be published in *Caves and Caving*.

References to previously published work should be given in the standard format used in *Cave and Karst Science*. In the text the statement referred to should be followed by the relevant author's name and date (and page number, if appropriate) in brackets. Thus: (Smith, 1969, p.42). All such references cited in the text should be given in full, in alphabetical order, at the end. Thus: Smith, D.E., 1969. The speleogenesis of the Cavern Hole. Bulletin Yorkshire Caving Assoc., Vol. 7. p.1-63. Books should be cited by the author, date, title, publisher and where published. Periodical titles should be abbreviated in standard style, or, where doubt exists, should be written out in full.

Acknowledgements: Anyone who has given a grant or helped with the investigation, or with the preparation of the article, should be acknowledged briefly. Contributors in universities and other institutions are reminded that grants towards the cost of publication may be available and they should make the appropriate enquiries as early as possible. Expedition budgets should include an element to help publication, and the editor should be informed at the time of submission.

Figures: Line diagrams and drawings must be in black ink on either clean white paper or card, or on tracing paper or such materials as Kodatrace. Anaemic grey ink and pencil will not reproduce! Illustrations should be designed to make maximum use of page space. Maps must have bar scales only. If photo-reduction is contemplated all lines and letters must be large and thick enough to allow for their reduction. Letters must be done by stencil, Letraset or similar methods, not handwritten. Diagrams should be numbered in sequences as figures, and referred to in the text, where necessary, by inserting (Fig. 1) etc. in brackets. A full list of figure captions should be submitted on a separate sheet.

Photographic plates are welcome. They must be good clear black and white prints, with sharp focus and not too much contrast; prints about 15 x 10 cm (6 x 4 inches) are best; if in doubt, a selection may be submitted. They should be numbered in sequence but not referred to in the text, except where essential and then after discussion with one of the Editors. A full list of plate captions, with photographer credits where relevant, should be submitted on a separate sheet.

Tables: These should not be included in the text but should be typed, or clearly handwritten, on separate sheets. They should be numbered in sequence, and a list of captions, if necessary, should be submitted on a separate sheet.

Approximate locations for tables, plates and figures should be marked in pencil in the manuscript margins.

Copyright: It is the author's responsibility to obtain written permission where necessary to reproduce all material submitted for publication.

Speleological expeditions have a moral obligation to produce reports (contractual in the case of recipients of awards from the Ghar Parau Foundation). These should be concise and cover the results of the expedition as soon as possible after the return from overseas, so that later expeditions are informed for their planning. Personal anecdotes should be kept to a minimum, but useful advice such as location of food supplies, medical services, etc. may be included, normally as a series of appendices.

Authors will be provided with 20 reprints of their own contribution, free of charge, for their own private use.

We prefer articles to be submitted on disk if possible, although paper copy is also acceptable. We can read most PC based word processing packages but if in doubt please consult one of the Editors. Apple Mac disks are accepted as a last resort!

If you have any problems regarding your material, please consult either of the Editors in advance of submission.

Cave and Karst Science

TRANSACTIONS OF THE BRITISH CAVE RESEARCH ASSOCIATION

Volume 23 Number 2 October 1996

Contents

Cover photo:
Measurement of radon daughter concentrations in Bounty Pot, Kirktau Massif, Russia.
Photo by John Gunn (see article by Clark Friend)
The photograph shows a Thompson and Nielson 'Radon Sniffer' and a digital barometer for simultaneous measurement of barometric pressure. The results are discussed in Cave Science 18(2), 1991, p. 85-87.

Editors: Dr. D. J. Lowe — British Geological Survey, Keyworth, Nottingham, NG12 5GG.
Prof. J. Gunn — Limestone Research Group, Department of Geographical & Environmental Sciences, The University of Huddersfield, Queensgate, Huddersfield, HD1 3DH.

Cave and Karst Science is published three times a year by the British Cave Research Association and is issued free to all paid up members of the Association.
The 1996 subscription rate to Cave and Karst Science is £16.00 per annum (postage paid).
Individual copies, back issues and details of annual subscriptions to Cave and Karst Science, and of Association membership, can be obtained from the BCRA General Administrator, 20 Woodland Avenue, Westonzoyland, Bridgwater TA7 OLG.
The Association's permanent address is: BCM BCRA, London WC1N 3XX.

© Copyright the British Cave Research Association, 1996. No material appearing in this publication may be reproduced in any other publication, used in advertising, stored in an electronic retrieval system, or otherwise used for commercial purposes without the prior written consent of the Association. All views expressed are those of the individual authors and do not necessarily represent the views of the Association unless this is expressly stated.
ISSN 1356-191X

EDITORIAL

John Gunn and David Lowe

Our previous editorials have focused on this journal and our policy concerning the types of material we wish to see appearing within it. Inevitably this could lead to a danger of repetition and accusations of "navel gazing". Hence, in this issue, we have decided to adopt a more global perspective to view cave and karst journals in general, although confining ourselves primarily to English language publications.

In Britain, apart from *Cave and Karst Science,* two journals publish a significant number of papers on aspects of cave and karst science. These are *the Proceedings of the University of Bristol Spelaeological Society* and *Studies in Speleology*, which is produced by the William Pengelly Cave Studies Trust. Both journals operate a system of peer review, and both include broadly similar material to that covered by *Cave and Karst Science*. It is inevitable then that the question will arise as to whether it would be better if there was a single cave and karst science journal, which might have a wider circulation and would be able to select the "best" papers for publication. As editors we have neither an "axe to grind" nor any control of policies; any such "merger" would have to be agreed by the councils or committees of the organisations concerned. However, it would be interesting to hear from anyone with strong and reasoned views either "for" or "against" such amalgamation.

The closest English language analogue to *Cave and Karst Science* (the *Transactions of the British Cave Research Association*) is *The NSS Bulletin* produced by the National Speleological Society of America. Until 1995 this was subtitled *Journal of Caves and Karst Studies* but from April 1996 [Volume 58, Number 1] the headings were switched so that *Journal of Caves and Karst Studies* is now the banner title with *The National Speleological Society Bulletin* as the sub-head. Like *Cave and Karst Science*, the *Journal of Caves and Karst Studies* publishes peer reviewed papers submitted by both domestic and overseas authors. During the late 1980s and early 1990s the production schedule of the *Bulletin* became quite erratic, as the editors struggled to attract good quality material. From time to time the situation with the former *Cave Science* (while not so severe due to the sterling efforts of Trevor Ford) presented some parallels here! Once again then, the question arises as to whether some form of amalgamation between *Cave and Karst Science* and the *Journal of Cave and karst Studies* would be beneficial to the international community of cave and karst scientists. Australasian cavers also have their own scientific journal, *Helictite*, the Journal of Australasian Speleological Research, which is published by the Speleological Research Council Limited, a non-profit limited liability company. Although the articles published therein have a more restricted geographical coverage than those in the two *"Cave and Karst"* journals, there is an area of commonality that could perhaps be explored.

Moving from journals produced, and sponsored by, national caving groups, we must also consider the *International Journal of Speleology,* which is sponsored by the International Speleological Union. Without any bias, we feel that it is fair to say that this journal has had a somewhat chequered history. It has never achieved its full potential, either in terms of circulation or reputation, when compared to the other journals mentioned above. Perhaps there is scope to consider a pooling of resources and effort to produce a high quality *"International Journal of Cave and Karst Science",* which would be circulated to all members of the British Cave Research Association and the National Speleological Society, and would also be sponsored and promoted by the International Union of Speleology, of which there are no individual members, and possibly also by the Speleological Research Council Limited or the Australian Speleological Federation.

Such a move would, of course, run counter to general trends in academic publishing, where new journal titles appear almost daily. In addition, some academics would argue that there is no advantage in "preaching to the converted", and that instead cave and karst scientists should aim to publish their work in subject-specific journals, thereby alerting a wider audience to the potential importance of the many aspects of caves and karst. These journals are generally perceived to have a higher "status" and to attract more academic kudos than journals designed specifically for those interested in caves and karst. Our view is that it should not be a case of "either/or" but of "both/and". It is important that the wider scientific community should be made aware that the study of caves and karst can provide important insight into a broad range of scientific problems, but it is equally important that the study of caves and karst be seen clearly as a serious scientific discipline in its own right. Clearly the establishment of a leading *"International Journal of Cave and Karst Science"*, equivalent in status to leading journals in the fields of archaeology, ecology, geology, geomorphology, hydrology and so on, would be one way to achieve this. However, there would be a great risk of alienating the wider community of "cavers", many of whom gain valuable information from the specifically cave and karst journals. Their support would be vital to the production of such an international journal, just as the support of the BCRA membership is an essential contribution to *Cave and Karst Science* in its present format.

We could go on to discuss many other excellent journals that publish articles on cave and karst science in languages other than English: Karstologia, Acta Carsologica and Studia Carsologica, to name but a few. However, that would lead us into areas of still greater complexity, as would consideration of the possibilities of publishing an electronic journal!

At this point we must emphasise that the content of this editorial has not been discussed with the Council of the British Cave Research Association. It is simply a collection of currently circulating ideas, brought together with the intention of stimulating debate on these issues amongst Council and membership of the BCRA, as well as other cave and karst scientists. Our own views and preferences may be immaterial and are, in any case, embryonic at this stage. However, our hope is that we will be informed by any ensuing debate, rather than being castigated for daring to describe such radical potential changes! Needless to say, our other aim in writing this editorial is, as usual, to encourage more people to submit papers to *Cave and Karst Science*. Without such contributions, the discussion is indeed "academic"!

CAVE AND KARST SCIENCE Vol. 23, No.2, October 1996
Transactions of the British Cave Research Association

Radon exposure during underground trips: a set of guidelines for caving and mine exploration in Britain

Clark R.L. FRIEND (Editor)
The National Caving Association, Monomark House, 27, Old Gloucester Street, London WX1N 3XX

Abstract: The natural processes that lead to the formation and distribution of the radioactive gas radon and its daughter products in caves and mines are examined, together with the ways that their presence can be recognised and measured. Problems presented to those who visit the caves and mines are discussed and their potential impact assessed. Against this broadly scientific background, those aspects of British law which relate to potential radon exposure underground are considered and the guidelines recommended by the National Caving Association are outlined.

INTRODUCTION

This paper is derived from a document prepared for the National Caving Association (NCA) by its specially-formed Radon Working Party (National Caving Association, 1996). The latter document was prepared in response to the recent enactment of a law concerning safety of young persons at Outdoor Activity Centres. Prior to the establishment of the Radon Working Party, only show caves and mines had resolved the requirements of compliance with health and safety legislation in respect of the known hazard from radon. There was also a need to have guidelines for recreational cavers, for whom there were no recommended limits of exposure to radon and whose activities fall outside the provisions of the Health and Safety at Work Act. Further, there was a need to obtain guidance for the owners of caves, mines and land, as there were many potential problems due to the commonly complex situations regarding the ownership of land, mineral rights and cave/mine entrances. There was also some doubt over the extent to which persons with the gift of access to caves were legally liable for what happens underground with respect to radiation.

To rectify this situation the Radon Working Party was given the remit to prepare a set of guidelines that would assist with setting up safe working procedures to enable compliance with health and safety legislation. The document was prepared in consultation with the Health and Safety Executive (HE), who enforce health and safety legislation, and the National Radiological Protection Board (NRPB), who provide advice on radiation matters. However, because there have been no tests in court, these guidelines cannot be absolutely foolproof. None-the-less, the document sets out (a) the responsibilities that recreational cavers have to themselves and each other, (b) a reasonable set of working conditions with a monitoring protocol for professional cavers, and also (c) provides advice to owners.

Research has demonstrated that the air in most caves and mines contains varying concentrations of the radioactive gas radon and its daughter products. Because it is a gas, the concentrations of radon and its radioactive daughter products are not constant at any one location, being controlled in a complex manner. Recognition of the presence of radon in caves and mines potentially used by outdoor centres has highlighted the need to consider the application of the Ionising Radiations Regulations 1985 (IRR), as they are places of work. Presently there is no indisputable proof that cavers' health has been affected adversely by exposure to radon and its daughters during caving activities. However, research in uranium mines world-wide has shown that a risk does exist and although this paper is written from a British perspective there is international interest in cave radon (Hyland and Gunn, 1995). Hence, it is hoped that the proposed guidelines will be of wider interest. Professionals and others in positions of responsibility are reminded that the information in the guidelines should be considered in conjunction with the requirements of the *Cave Instructor Certificate* and the *Local Cave and Mine Leader Assessment Scheme*, both administered by the NCA.

RADON: GENERAL INFORMATION

Radon (Rn) is the heaviest naturally occurring gaseous element, and is part of the "Inert Gas" group of the Periodic Table. There are no stable isotopes of radon and of its 20 known radioactive isotopes, three are formed naturally by the radioactive decay of certain isotopes of uranium and thorium. Isotopes are two or more atoms of the same element that contain different numbers of neutrons, hence giving them different atomic weights, but the same number of protons, giving them the same atomic number. Different isotopes, for example radon's Rn 220 and Rn 222, have identical chemical properties.

Uranium (U) and thorium (Th), have long radioactive half-lives and have been present in the earth's crust since the earth was formed. They are found in ores such as pitchblende or in monazite and in granites, associated with other minerals. Erosion of the earth's crust gives rise to trace levels of uranium and thorium in sea water, which thus contaminates any rock that is formed through marine sedimentary processes. In addition, some plants and animals are capable of concentrating some elements, including uranium, in their bodies. If, for example, uranium is concentrated in this way, it gives rise to elevated levels of uranium in rock derived from the plant or animal bodies. Thus limestone will contain background levels of uranium and thorium that can increase by substantial amounts locally. Other sedimentary rocks, such the Northampton Sands, have moderate trace levels of uranium due to other concentrating processes acting during deposition.

The radioactive decay of the uranium and thorium series parents occurs by the emission of either an alpha particle (decays between isotopes in the same row in Fig. 1) or a beta particle (decays between isotopes in the same column in Fig. 1), together with the emission of various gamma particles or rays. Both the ejected particle(s) and the decayed atom also hold some energy, normally in the form of momentum, resulting from the decay. This energy is normally measured in the units known historically as Mega electron Volts (MeV); the larger the value the greater the energy. All three particles or rays are known as ionising radiation, as they all have the property of ionising molecules as they pass through the air. Alpha particles consist of two protons and two neutrons that are ejected from the nucleus of the decaying atom, leaving the atom with two spare electrons. A beta particle is an electron ejected from the nucleus when a neutron transforms into a proton, leaving the atom with a positive charge. Fig. 1 shows the decay paths and identifies the subsequent "daughter products" of the radon isotopes that eventually decay into stable isotopes of lead (chemical symbol Pb), together with their half lives and energy of alpha particles. Fig. 1 has been simplified to ignore a number of different decay paths that occur at trivial levels. Given a fresh quantity of Rn 222, the daughters will grow in number with passing time as shown in Fig. 2. Under normal circumstances, radon will have reached some level of equilibrium with its daughters that is dependant on many factors; some of which can vary on short time scales.

Uranium Series (radon)

U 238 4.5E9 y	Th 234 24 d							
	Pa 234 6.7 h							
	U 234 2.5E5 y	Th 230 7.7E4 y	Ra 226 1600 y	Rn 222 3.82 d 5.49 MeV	Po 218 3.05 m 6.00 MeV	Pb 214 26.8 m		
						Bi 214 19.9 m		
	Alpha u particle			Beta p particle		Po 214 2E-4 s 7.69 MeV	Pb 210 22.3 y	
							Bi 210 5.0 d	
							Po 210 138 d	Pb 206 stable

Thorium Series (thoron)

Th 232 1.4E1 0 y	Ra 228 5.8y					
	Ac 228 6.1 h					
	Th 228 1.9y	Ra 244 3.7 d	Rn 220 55.6 s	Po 216 0.15 s	Pb 212 10.6 h	
					Bi 212 60.6 m	Tl 208 3.07 m
					Po 212 3E-7 s	Pb 208 stable

Actinium Series (Actinon)

U 235 7.0E8 y	Th 231 25 h						
	Pa 231 3300 y	Ac 227 22 y	Fr 223 22 m				
		Th 227 19 d	Ra 223 11 d	Rn 219 4.0 s	Po 215 2 E-3 s	Pb 211 36 m	
						Bi 211 2.1 m	Tl 207 4.8 m
							Pb 207 stable

Figure 1. A Simplified presentation of the Uranium, Thorium and Actinium Series.
Data from International Commission on Radiological Protection 1983. Trivial Branches have been omitted for clarity and only principal alpha energies included.

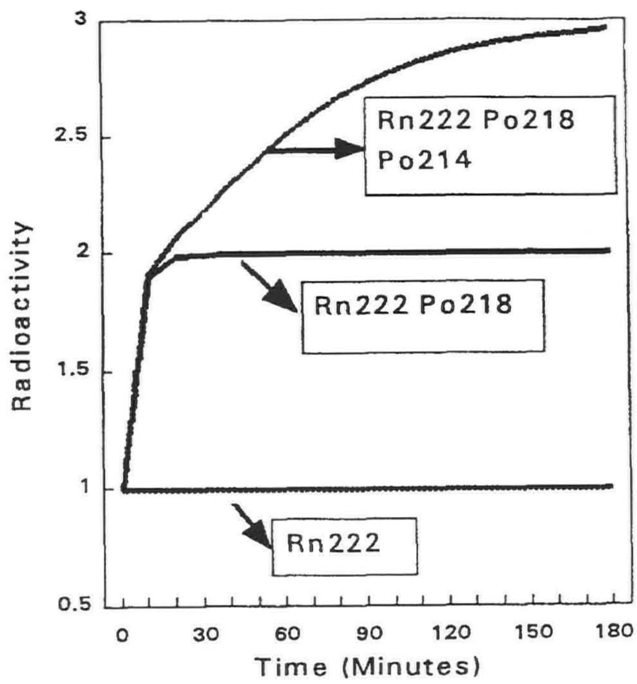


Figure 2. Rate of In Growth of Radioactivity from Daughters of Rn 222.

Measurement of radioactivity is based on measuring these particles. For radon and its daughters, there are three different methods, each with variations. The earliest method devised was based on the principle that the ionising particle hits molecules in the air causing them to become ion pairs. Friedmann (1983) developed an instrument that measures the current flow from a collector held at a negative voltage to its chamber caused by the ion pairs separating under the potential difference and reaching the collector. Thus an indirect measurement of the number of alpha particles may be made. Because of the energy associated with the particle, many ion pairs are formed, hence amplifying the size of the signal. This ionisation method will measure all ionising radiation within the chamber, which includes the alpha and beta particles from radon together with its daughters, plus any background gamma particles. A variation of the ionisation method is to take an air sample, either directly or by absorbing the radon onto charcoal, and subsequently measure it. Whilst this can reduce the background gamma problem, it also loses the contribution from the daughters.

The second method is based on measuring the ionising particles themselves by means of their interaction with a solid state electronic detector causing a current flow. Whilst gamma particles will interact throughout the volume of the detector, alpha and beta particles are measured by their interaction in the surface layer of the detector. This technique can be used to look directly at the radiation in the air or, more usually, by looking at the level of radiation collected on a filter. Although the radon will pass through a filter, the daughters are very efficiently collected due to their charge. There are two principle variants. One measures the build up of radioactivity on the filter directly as air is sucked through it, as employed in some direct measuring devices such as the Thompson and Nielson 'Radon Sniffer'. The other, proposed by Kusnetz (1956), looks at the filter paper after a period of time following the collection of the daughters. Thomas (1972) subsequently proposed a technique whereby counting the level of alpha particles after several periods of time, enables the calculation of the concentrations of the separate daughters. Some solid state detectors, known as spectrometers, are able to measure the energy of the particle as well, and are thus able to measure the individual isotopes directly.

The third method is to measure the impact of the ionising particles on a sensitive material that can subsequently be developed (similar to a photograph) to display the individual impacts. Several types of plastic have been produced that, on etching, will display the individual impact of alpha particles. These have been incorporated by Camplin et al (1988) into a simple passive detector and by the National Radiological Protection Board (Bartlett, 1987) into a personal dosemeter. The sensitivity of this material is relatively low and thus these detectors are normally exposed to air for a period of several days or longer. Hence they are used to record concentrations averaged over these longer periods, rather than the minutes that the other two methods can achieve. The other feature of these devices is that they record alpha particles from radon and its daughters that are ejected in close proximity to the surface of the detector. As the NRPB device has been designed such that only radon atoms can diffuse through the detector housing, this devise records the level of radon in the air plus the growth of daughters that takes place within the housing. Deliberately, it does not measure the level of daughters present outside the housing. (This results in an over estimate, since the external concentration of daughters is unlikely to reach the equilibrium levels that will occur within the housing.)

The measurement of radon and its daughters is a complex subject and the National Council on Radiation Protection and Measurement of the USA has produced a detailed publication on the topic (NCRP, 1988). For cavers, apart from purchasing or manufacturing instruments, perhaps the simplest means for making long term measurements is to use the same material as the Camplin et al (1988) detectors. This can be obtained from Track Analysis Systems Ltd (1996). Hyland (1995) measured the concentration of radon and daughters in a number of British caves and his results are summarised by region in Table 1.

Because radon is an inert gas, it can diffuse away from its birth place relatively rapidly and, if near enough, reach the boundary of the grain of rock within which it was born. Given small enough distances, the radon atom can thus travel along a rock grain boundary to a pore and thence to a crack and eventually into the free atmosphere. The proportion of atoms escaping in this manner is also controlled by the half-life of the particular radon isotope. Hence, the 3.9 second half-life of Rn 219 effectively means that none is released. (Rn 219 was called actinon by its discoverers, as it was erroneously thought to be derived from the element actinium, rather than from U 235.) The 56 second half-life of Rn 220 means that very little escapes from the rock. (Rn 220 was called thoron by its discoverers, as it was associated with the parent element thorium.) In the United Kingdom, it has become acceptable to ignore the impact of Rn 220 as a counterbalancing assumption to the many other pessimistic assumptions made in assessing the impact of radon. Thus the dominant radioactive species of concern is Rn 222 or radon, whose name derives from its

Region	Radon Concentration Bq m^{-3}			
	Minimum	Maximum	Mean	Standard Deviation
Portland	10	974	454	326
North Pennines	14	27136	1115	2089
Mendip Hills	99	3621	1129	1057
South Wales	127	19968	2561	2773
Peak District	9	46080	8868	10724

Table 1. Examples of radon concentrations (Bq m^{-3}) measured in a number of caves in various parts of the UK in the period 1991-1992. (Data from Hyland, 1995).

radioactive parent, radium. The rate of release of radon is therefore dependant not only upon the concentration of uranium in the rock but also the surface area of the rock. For example, highly fractured limestone will expose more surface area than limestone with few fractures. Dust made by mining operations or eroded small particles of limestone will expose even larger surface areas and Bottrell (1991) has postulated that they might contribute significantly to underground radon levels.

Once released into the atmosphere, radon is subject to a number of influences as it decays progressively through its daughters. The most significant influence is air movement. The concentration of radon in the open is quite small due to the efficient mixing caused by air currents. Ventilation is a powerful means of reducing radon levels and has been deployed effectively in show caves and houses to reduce radon concentrations. However, in enclosed spaces the efficiency of air currents can be limited and where elevated carbon dioxide levels are known to persist, higher concentrations of radon could build up. A second significant factor is air pressure variation. A substantial pressure drop caused by a deep low pressure weather system passing across the country will cause air to flow out of the cracks, carrying radon with it. Hence the radon concentration will be elevated for a period linked to the time taken for the low pressure to pass. The third influence relates to differences in air temperatures. In winter, air outside a cave or mine that is relatively colder than the air inside will tend to flow inwards, hence "pushing" radon back into cracks. In the summer, with the reversal of relative temperatures, radon will be drawn out of the cracks. Elevated summer radon concentrations compared to those during the winter have been measured in a number of caves by Middleton et al (1991) and Hyland (1995). A potentially related effect may be due to a reduction of radon's solubility in water with increasing water temperature during the summer months.

Another influence is the charge left on the decayed atom, Po 218, an isotope of polonium, is formed with an initially excess negative charge. Disregarding chemical reactivity, by which the polonium atom could be transformed into polonium oxide, the charged species will be attracted to a surface. Toohey et al (1984) measured the plate-out of daughters on surfaces. However, the attracting surface could conceivably be an airborne particle, and the combination of decayed atom and airborne particle is referred to as attached. A parameter called the unattached fraction (the fraction of daughters not attached to particles) is significant in determining the impact of the daughters. Larger attached particles are subject to a settling out process. Further decay of Po 218 to Pb 214 could result in the Pb 214 atom retaining sufficient energy to detach itself from the particle and become unattached. The Pb 214 atom would then be subject to a similar process. Both attached particles and unattached atoms of the daughters can be removed by mechanisms such as scrubbing by water sprays in active cave or mine systems. Filtration is also effective in removing daughter products. These influences thus create a ratio of radon to daughters that is not dependent solely upon the radioactive decay of the isotopes.

Measurement of radiological impact

The SI unit of a Bequerel (Bq) is defined as the decay of one atom in one second. Normally the decay rate is measured by detecting the particle or particles emitted. There is a one to one relationship for alpha particle emissions from radon and its daughters and thus measurement of the quantity of radon and daughters present can be made by counting the number of alpha particles emitted in a given time. More conveniently for impact calculations, concentrations of radon and daughters in air can be measured and denoted in terms of Bq m^{-3}. Measuring systems using solid state detectors or sensitive material are normally calibrated in units of Bq. However, given the early use of measuring systems based on the ionisation principle, a special concentration unit called the Working Level (WL) was devised by the mining industry. This is based on a measurement of the total alpha ionisation energy associated with radon and its daughters decaying per unit volume of air. As one of the simplifying assumptions, in the United Kingdom, the definition was further narrowed to include only the alpha particles from Rn 222, Po 218 and Pb 214. This is based on the low level of thoron and the assumption that, given its relatively long half-life (22 years), the impact of Pb 210 can be ignored compared to that of its three predecessors. Considering the potential variation in the ratio of radon and daughters, due to factors discussed above, the Working Level was a pragmatic unit suitable for use in mines and thus has no simple relationship with the SI unit of Bq m^{-3}. A relationship can be expressed by use of another parameter, called the equilibrium constant. This is the ratio of the actual measured alpha activity level to that from the same quantity of radon that is in radioactive equilibrium with its daughters (see NCRP 1988). Care is needed to avoid confusion of this equilibrium constant with another frequently referred to concept, known as the Equilibrium Equivalent (Radon) Concentration (EEC). The EEC is the concentration of radon, in radioactive equilibrium with its daughters, that emits the same potential alpha energy per unit volume as that in the measured mixture.

The potential impact of exposure to radiation is measured in terms of time integrated exposures or "dose". The dose of radiation received by a person is related to the concentration of radon and daughters and the length of time for which the air is breathed, that is Bq m^{-3} h. The SI unit of dose is the Sievert (Sv). Account is also taken of the difference in nature of the ionising particles that are delivering the dose. Dosimetric calculations can be carried out to derive a relationship between the time integrated concentration and dose. These calculations are based on estimating energy deposited by the particle into the cells of organs in the body, together with modifying factors to account for the potential of the particle to cause damage. In the case of radon and its daughters, the organ of interest is the lungs. Due to its inert nature radon will penetrate with breaths of air into the farthest reaches of the lungs. The extent to which the daughters penetrate into the lungs depends upon the unattached fraction of daughters. The equilibrium constant characterises the amount of each daughter present within the lungs. Based on this information, it is possible to calculate the dose of radiation given by radon and its daughters to the lungs from a given concentration, composition and unattached fraction. It should be noted that the precision implied by this description is usually inappropriate, as most measurements do not determine the equilibrium constant, let alone the unattached fraction. In any case, the imprecision of many measurements is larger than the effect of neglecting the unattached fraction or assuming a typical value for the equilibrium constant.

Table 2. Dose limits for classes of persons as defined in the Ionising Radiations Regulations.

	mSv	WLh (precise)	WLh (practical)[1]
Employee Dose Limit	50[2]	816	680
Any Other Person Dose Limit	5	82	68
ICRP recommendation for any other person	1	34	28

Note: [1] The practical value is reduced from the precise value to take account of associated ionising radiation from the decay of radon, radon daughter products and other associated radioactive material that is not detected by instruments measuring Working Levels. The data are taken from the Health and Safety Commission (1988) using International Commission for Radiological Protection (1981, 1991 and 1993). The ICRP values are based on workers rather than general public, since the vast majority of those entering underground voids are adults of working age.

[2]As soon as the level of exposure reaches 15 mSv reassessment of procedures must take place.

An ionising particle colliding with a cell will deposit energy into the cell causing damage. There are three potential outcomes of that damage - the cell either dies, repairs itself or becomes cancerous. The major risk from radon is therefore one of developing lung cancer. Information is available to relate dose to risk of cancer. The International Commission on Radiological Protection (ICRP) has made recommendations on the limit to which persons should be exposed to radiation, and these have been incorporated into United Kingdom legislation (Table 2). They are based upon typical doses received by workers exposed occupationally to radiation, being equivalent to a risk level similar to that found in hazardous occupations recognised as having high safety standards (ICRP, 1977). Since then, further evidence has shown that the risks were underestimated, and a reduction in dose limits (Table 2) has been advised by the ICRP (1991). At present this has not been incorporated into United Kingdom legislation.

An alternative approach to risk, based upon epidemiological studies of uranium miners, has been studied by the ICRP. The equivalent unit of dose used in the mining industry is the Working Level Month (WLM), which is the exposure of a person to 1 WL for a period of 170 hours (a typical time a person would spend exposed underground in one month). The ICRP (1993) have more recently advised using the results of these studies that were reported in Working Level units. One of the problems in quoting data is that the basis for the conversion from measured units to dose has changed over the past twenty years. Using an equilibrium factor of 0.5, Kendall et al (1994) showed that exposure to 170 WLh is equivalent to a time integrated exposure of 1.26×10^6 Bq $m^{-3}h$ or 144 Bq m^{-3} y , that is, a person breathing air with radon concentration of 144 Bq m^{-3} continuously for a whole year. **The data quoted in this document should not be used to derive relationships between units without consulting the source documentation.**

RISKS FROM RADON

To place these figures in context, the average concentration of radon alone indoors in the United Kingdom is about 20 Bq m^{-3} (O'Riordan, 1990). Taking into account a wide range of factors, O'Riordan (1990) calculated this to correspond to an annual dose of about 1 mSv to a person living in the house. He also reported that the range of radon concentrations found in houses across the country varies between 5 and 5000 Bq m^{-3}. It has also been calculated that the annual average dose to a person living in Cornwall due to radon is about 6.3 mSv, compared to 1.2 mSv for a person living in London (Metters, 1992).

Because some houses have high radon concentrations and thus expose persons living in them to substantial doses of radiation, there have been a number of recommended action levels. The principle is that if an action level is found to be exceeded, steps should be taken to reduce the radon concentration. Although the European Union (as the Community of European Countries) recommended action levels of 400 Bq m^{-3} for existing houses and 200 Bq m^{-3} for new houses (CEC, 1990), the NRPB in 1990 recommended an action level of 200 Bq m^{-3} for all houses. This 200 Bq m^{-3} action level was calculated to correspond to an annual dose of about 10 mSv to a person living in the house (O'Riordan, 1990), which is consistent with the latest ICRP recommendations (ICRP, 1993).

Whilst the NCA do not wish to trivialise the risks involved, there is a tendency for the media to sensationalise the risks from any source of radiation, including radon. In risk consequence terms, an average radon concentration of 20 Bq m^{-3} in a home over a person's lifetime is estimated to give rise to a lifetime risk of lung cancer equivalent to 0.3% (ICRP, 1987; O'Riordan, 1990). In other words, there is a probability of 3 in 1000 that an individual would die from the specific cause of lung cancer due to the lifetime exposure to radon in their home. It should be noted that the prevailing risk of lung cancer in the community is about 6% and is equivalent to a lifetime exposure to radon at a concentration around 400 Bq m^{-3} within the home. Additionally, current medical opinion is that lifetime smoking increases the risk of dying from lung cancer tenfold. It is considered that exposure to radon at 200 WLh per year for life doubles the risk of dying from lung cancer. Note that estimates for smokers exposed to radon suggest a multiplicative, not additional, risk of contracting lung cancer (Table 3), though some opinion suggests that there may be more variation (Doll, 1992; Peto and Darby, 1994).

Table 3. Statistical data for health risks associated with smoking and exposure to radon (from Metters, 1992).

	Average in UK houses (20 Bq m^{-3})	House at radon action level (200 Bq m^{-3})
Lifetime risk of lung cancer due to radon in general population	3 in 1000	1 in 100
Lifetime risk of lung cancer due to radon in smokers	3 in 100	1 in 10

Metters (1992) points out that areas with acknowledged high radon concentrations such as Cornwall and Devon have lower death rates from cancer than London or Manchester, which are areas of low radon concentration, so it is clear that there is room for scientific debate. Other work concludes that this kind of geographical study has shortcomings, such that it does not contribute to a better understanding of the quantitative risks from indoor radon (eg. Stidley and Samet, 1993). There is also speculation over whether the application of risk factors for mines (from which the exposure limits in legislation were derived) can be transferred fairly to the conditions found in most caves. Air in mines tends to have other particles (dust, fumes) that are themselves harmful and thus add to the risks, though some studies have corrected for these risks.

The persons engaged upon caving activities as a recreation do so *willingly* and as such accept the associated hazards of their chosen sport - which are many. Some hazards are immediately obvious (for example, the risk of drowning), whilst others are intangible (e.g. the risk of contracting Weil's Disease). The risks involved in entering an underground void relate to three main sources. First, the ability level of the individual, e.g. climbing skills and/or use of technical equipment. Second, the hazards of the void itself; for example the danger of rockfalls. Third, aspects of the cave/mine environment, such as the possibilities of flooding or bad air. The NCA recognises radon exposure as one potential underground hazard, and will endeavour to ensure that members are kept updated on the nature and extent of the risks as new information becomes available. Deaths from radon-related causes are only likely to become apparent some years after persons have given up active caving. However, at present, the NCA is not aware of any information to suggest that there is a high death rate from lung cancer among retired cavers.

It is possible to compute the approximate risk of death from exposure to radon resulting from a single caving trip. Although radon concentrations vary considerably, a typical caving trip of around 4 hours duration would lead to a person receiving a dose of around 0.1 mSv from exposure to radon (Table 1). Using a risk relationship of 0.056 chance of death per Sv (ICRP, 1993), the typical trip would create a chance of the order of 6 in 1,000,000 of dying subsequently from lung cancer induced by this exposure. In comparison, it is possible to make an order of magnitude estimate of the actual risk from caving. Although there are no census figures it is estimated that there are about 10,000 active cavers in the UK. Given the considerable variability of the frequency that this population actually goes caving, an average of 10 trips per year is probably optimistic. The actual number of deaths of persons on caving trips in the UK and Eire between 1982-1993 was 31, of which 6 were of divers (British Cave Rescue Council, 1994). Because cave diving involves a very small number of persons who expose themselves to hazards quite separate from those faced by ordinary cavers, the deaths attributable to diving have been ignored in this calculation. As deaths are about 2.5 per year this gives an estimated risk of death for a recreational caver from an accident in a cave during one caving trip of around 25 in 1,000,000.

Exposure to **all** radiation should be **as limited as is possible** but, as with anything, there is a risk. It is the recommendation of the NCA that, wherever possible, individuals should:

- attempt to find out as much as possible regarding likely exposure to radiation from radon on each trip undertaken;
- use known values to make rough calculations of the potential cumulative dose during a year;
- attempt to keep exposure within NRPB recommended values.

DEFINITIONS

Definitions of certain words and phrases were adopted to facilitate understanding and help clarify legal matters. The phrase underground void was adopted to cover any cave, mine, show cave, show mine or other underground void, whether natural or artificial, into which a person could enter. Two broad groups of persons were identified and sub-groups defined as follows:

*1. **Recreational***

i) Cavers and Mine Explorers being persons who explore underground voids as a pastime. These persons have no contract of employment underground for the activity they are engaged upon. This category includes persons with specialist local knowledge of a cave or mine who may be requested, on a voluntary basis, to take a recreational party through that cave or mine.

ii) Voluntary Group Leader being an adult leading or instructing a group, normally on behalf of a charitable organisation such as the Scouts or youth group, who is in a position of acknowledged authority. There is no contract of employment and no financial reward.

iii) Members of Instructed or Led Parties, ie. persons under the charge of an Instructor or Voluntary Group Leader.

*2. **Professional***

i) Instructors, being persons who provide a service of instructing others in underground exploration techniques and practices in any type of underground void. Such persons are normally either self-employed or are employed by, for example, an Outdoor Education Centre.

ii) Research scientists, being persons employed, engaged, commissioned or directed to carry out scientific investigations in all types of underground void.

iii) Show Cave/Show Mine Guides and other persons being employees of a Show Cave or Mine Owner who either maintain the Show Cave or Mine, or guide and assist visiting members of the General Public who were excluded from further consideration.

iv) Employers, being those who employ the persons listed above, or those who are self-employed.

Four groups of owners of underground voids and land were defined to facilitate the explanation of legal duties. There are two distinct types of persons, corporate bodies, lessees, licensees or other associations within these definitions. The first type is concerned with the underground void, whilst the second type is concerned with the land surface. Clearly they are not mutually exclusive, but commonly are in practice. The key feature of either type is that they have the legal right to control access either within the underground void or across the land surface up to the entrance of the underground void. The term "Owner" was used to include all other alternatives who may hold the legal right of access. Whilst ownership of a working mine is known, ownership of a natural cave or abandoned mine is commonly not clear and could reside with the owner of either the mineral rights or the land surface. These groups are:

i) Private Void Owners - ie. owners of a natural cave or abandoned mine or other underground void or part thereof, which does not form part of the owner's business (eg. an owner of a private estate).

ii) Private Land Owners - ie. owners who have control of access of the land surface that is not part of the owner's business, exclusive of their legal relationship to the underground void (eg. a householder)

iii) Business Owners of Voids - ie. owners of a natural cave or mine (working or abandoned) or other underground void, or part thereof, which does form part of their business (eg. Show Caves and Mines).

iv) Business Land Owners - ie. owners who have control of access of the land surface as part of their business, exclusive of their legal relationship to the underground void (eg. a farmer).

Miners, Mine Managers and Mine Operators were excluded from the document since their legal duties are already well understood. In addition, members of the general public were excluded on the basis that those who might go caving were caught by one of the defined groups and did not need separate consideration.

UNITED KINGDOM LEGISLATION

Five pieces of United Kingdom legislation were considered by the Working Party; namely the Occupiers Liability Acts of 1957 and 1984, the Health and Safety at Work Act of 1974, the Mines and Quarries Act of 1954 and the Ionising Radiations Regulations of 1985. The Occupiers Liability Acts creates a duty of care on the land owner towards visitors, except those who willingly accept the risks. The Health and Safety at Work (HaSW) Act creates a range of duties primarily on employers to safe guard their employees and others from risks created by the work situation. The Mines and Quarries Act creates a duty to secure entrances to abandoned mines. The Ionising Radiations Regulations create a range of specific duties on employers whose work involves exposure to ionising radiations. Because of the nature of radon, the regulations were carefully worded so as to also include work activities that took place in an atmosphere that contained radon above a fairly low specified concentration. Further details of this legislation can be found in the NCA document (1996).

GUIDELINES FOR MINIMISING EXPOSURE TO RADON

These guidelines are designed to assist organisations and individuals concerned with and engaged upon underground activities, to adopt practices that will enhance awareness of the problems of and controlling exposure to radon and its daughter products. These guidelines are in amplification of the normal codes of practice for conducting parties underground prepared by NCA through its various sub-committees. Also, they provide a framework to which the Cave Instructor Certificate and Local Cave and Mine Leadership Assessment Scheme can refer.

These guidelines **do not apply to underground voids where radon concentrations are abnormally high**, such as certain metalliferous mines. In such places where only respiratory protection can provide adequate protection from radon, specific advice must be sought from a competent person before entering the underground void.

These guidelines take account of the fact that any health risks arising from exposure to radon are cumulative and so any unnecessary exposure should be avoided. NRPB have proposed for recreational caving a time integrated dose limit of 106 Bq m^{-3} h in any year (Kendall, 1995).

A. Recreational

1. Cavers and Mine Explorers

You should use the information available to plan trips that will minimise your own exposure to radon and avoid, wherever possible, taking others into underground voids that have high radon concentrations. If your cumulative time underground is likely to be in hundreds of hours, you should seriously consider carrying a dosemeter. An alternative, but less effective, way of estimating exposure is to use available data on radon concentrations in the voids you visit to obtain an approximate cumulative figure. You should comply with any rules that a land or void owner imposes as a condition of entry to an underground void. Such rules could be needed by the owner to protect their legal interests and may contain references to radon.

2. Voluntary Group Leaders

Voluntary Group Leaders are not covered by the HaSW Act but may be held to have a common law duty to members of their party. As part of their leadership function, leaders should advise members of their party about the hazard of radon and give them the best available information about the levels of risk associated with the proposed trip. If underground voids known or suspected to have high concentrations of radon are avoided, it should be sufficient to advise the party that radon is present, but the anticipated levels are such that the estimated risk is not out of proportion to other risks. The party should also be made aware of the cumulative risk. The Leader should ensure that the party complies with any rules made by the Land or Void Owner (see above).

3. Members of Instructed or Led Parties

Professional Instructors have duties under the HaSW Act to ensure that members of their parties are provided with adequate information, instruction and supervision. The principal measure to minimise exposure will often be the choice of suitable underground voids or parts of underground void. If you are a member of such a party you should follow the leader's instructions relating to the parts of the void to be visited. Although Voluntary Group Leaders are not covered by the HaSW Act, they may be held to have a common law duty to members of their party, so you, as a party member, should follow their instructions and advice on the areas to visit.

B. Professional

4. Employers

You are required to comply with the HaSW Act and the IRR. (Further details can be found in the NCA document (1996)).

5. Professional Instructors

Professional Caving Instructors have legal duties under the IRR. In the first instance you should either, if self employed, appoint a Radiation Protection Advisor (RPA), or if employed, be briefed by the RPA or Radiological Protection Supervisor appointed by your employer. (Further details can be found in the NCA (1996) document). In addition, you should comply with any conditions for entry to the underground void that the land or void owner considers to be necessary to protect their legal interest, including any duties under Section 4 of the HaSW Act

6. Research Scientists

The work of research scientists is potentially so variable that it is not practical to give general guidance. The comments in the HaSW Act will apply according to whether you are an employee or self-employed. You should also comply with any conditions for entry to the underground void that the Land or Void Owner considers to be necessary to protect their legal interest, including any duties under Section 4 of the HaSW Act.

C. Owners

7. Private Void Owners

You have no duties under either the HaSW Act or the IRR because you are not engaged in a business. Duties under the Occupiers Liability Acts may be discharged by an adequate sign. However, if you "own" a mine, you have a duty under the Mines and Quarries Act to maintain the entrance(s) secure so as to prevent a person from entering accidentally.

8. Private Land Owners

You have no duties under either the HaSW Act or the IRR because you are not engaged in a business. Duties under the Occupiers Liability Acts may be discharged by an adequate sign. However, if you have a mine entrance on your land, you may have a duty under the Mines and Quarries Act to maintain it secure so as to prevent a person from entering accidentally.

9. Business Void Owners

Most Business Void Owners are Show Cave or Mine owners. You have duties under the HaSW Act and almost certainly under the IRR because of the foreseeable presence of radon, at least insofar as access to that part of the underground void that is defined as the Show Cave or Mine is concerned. You should have already appointed a RPA. Other business Void Owners should determine the extent to which the IRR apply and hence whether they need an RPA. (Further details can be found in the NCA (1996) document). Duties under the Occupiers Liability Acts may be discharged by an adequate sign. However, if you "own" a mine, you have a duty under the Mines and Quarries Act to maintain the entrance(s) secure so as to prevent a person from entering accidentally.

10. Business Land Owners

Although you have duties under Section 4 of the HaSW Act to persons at work entering an underground void through an entrance on your land, these duties are limited to taking such measures as are reasonable to take. The principal duty to comply with the IRR lies with the employer of those persons (or self-employed person) who are entering the void and who should have taken advice from their RPA. Your duties in respect of the radon hazard are unlikely to extend beyond providing relevant information to the employer. This duty is one of providing such information that you hold on the hazard from radon to the person who seeks that information. Duties under the Occupiers Liability Acts may be discharged by an adequate sign.

REFERENCES

Bartlett, D T and Bird, T V, 1987. *Technical Specification of the NRPB Radon Personal Dosemeter*. NRPB-R208, 7pp.

Bottrell, S, 1991. Radon Production and Release from Cave Sediments. *Cave Science*, 18, 79-82.

British Cave Rescue Council, 1994. British Cave Rescue Council Incident Report, 1993. *Descent*, 118, 27.

Camplin, G C, Henshaw, D L, Lock, S and Simmons, Z, 1988. A National Survey of Background Alpha Particle Radioactivity. *Phys. Educ.*, 23, 212-217.

Community of European Countries, 1990. Commission recommendation of 21 February 1990 on the protection of the public against indoor exposure to radon. *Official Journal of European Community*, L80, 26-28.

Doll, R, 1992. Risks from Radon. *Radiation Protection Dosimetry*, 42, 149-153.

Friedmann, H, 1983. A Portable Radonmeter. *Radiation Protection Dosimetry*, 4, 199-122.

Health and Safety Commission, 1988. *Exposure to Radon: The Ionising Radiations Regulations 1985, Approved Code of Practice - Part 3.* HMSO COP 23.

Hyland, R Q T, 1995. Spatial and temporal variations of radon and radon daughter concentrations within limestone caves. Unpubl. PhD thesis, University of Huddersfield.

Hyland, R Q T and Gunn, J, 1994. International comparison of cave radon concentrations identifying the potential alpha radiation risk to British cave users. *Health Physics*, 67, 176-9.

International Commission on Radiological Protection (ICRP), 1977. Recommendations of the International Commission on Radiological Protection. *ICRP 26*, Annals of the ICRP, 1, No. 3.

International Commission on Radiological Protection (ICRP), 1981. Limits for inhalation of radon daughters by workers. *ICRP 32*, Annals of the ICRP, 6, No. 1.

International Commission on Radiological Protection (ICRP), 1983. Radionuclide Transformations - Energy and Intensity of Emissions. *ICRP 38*, Annals of the ICRP, 11-13.

International Commission on Radiological Protection (ICRP), 1987. Lung cancer risks from indoor exposures to radon daughters. *ICRP 50*, Annals of the ICRP, 17, No. 1.

International Commission on Radiological Protection (ICRP), 1991. 1990 recommendations of the International Commission on Radiological Protection. *ICRP 60*, Annals of the ICRP, 21, Nos. 1-3.

International Commission on Radiological Protection (ICRP), 1993. Protection against Radon 222 at home and at work. *ICRP 65*, Annals of the ICRP, 23, No. 2.

Kendall, G M, 1995. Limiting radon exposure in caves and abandoned mines. *Radiation Protection Bulletin*, 165, 16-17.

Kendall, G M, Miles, J C H, Cliff, K D, Green, B M R, Muirhead, C R, Dixon, D W, Lomas, P R and Goodridge, S M, 1994. *Exposure to radon in UK Dwellings*. NRPB-R272, 43pp.

Kusnetz, H L, 1956. Radon daughters in Mine Atmospheres: A Field Method for Determining Concentrations. *American Industrial Hygiene Association Quarterly*, 17, 85-88.

Middleton, T, Gunn, J, Fletcher, S, and Prime, D, 1991. Radon Daughter Concentrations in Giant's Hole, Derbyshire. *Cave Science*, 18, 67-70.

Metters, J, 1992. Setting radon in context. *Radiation Protection Dosimetry*, 42, 159-164.

National Caving Association, 1996. *Radon Underground*. ISBN 0 9525520 2 7, 24pp.

National Council on Radiological Protection and Measurement (NCRP), 1988. *Measurement of Radon and Radon Daughters in Air*. NCRP Report No. 97.

O'Riordan, M C, 1990. Human Exposure to Radon in Homes: Recommendations for the practical application of the Board's statement. *Documents of the NRPB*, 1, 17-32.

Peto, J and Darby, S, 1994. Radon risk reassessed. *Nature*, 368, 97-98.

Stidley, C A and Samet, J M, 1993. A review of ecologic studies of lung cancer and indoor radon. *Health Physics*, 65, 234-251.

Track Analysis Systems Ltd (TASL) 1996. *TASTRAK Solid State Nuclear Track Detector*. TASL, H H Wills Physics Laboratory, Tyndall Avenue, Bristol, BS8 1TL.

Thomas, J W, 1972. Measurement of Radon Daughters in Air. *Health Physics*, 23, 783-789.

Toohey, R E, Essling, M A, Rundo, J and Wang Hengde, 1984. Measurements of the Deposition Rates of radon Daughters on Indoor Surfaces. *Radiation Protection Dosimetry*, 7, 143-146.

CAVE AND KARST SCIENCE Vol. 23, No.2, October 1996
Transactions of the British Cave Research Association

Tafoni (pseudokarst features) in the Maltese Islands

Chris O. HUNT
Department of Geographical and Environmental Sciences, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK

Abstract: In the Maltese Islands, small to medium sized shallow blind bulbous cave-like features and elongate notches are common near the base of outcrops of the Upper Coralline Limestone, and are also found on steeply sloping outcrops of other limestone formations. These features are identified as tafoni. Theories of the formation of tafoni and similar features are reviewed. It is suggested that, in the Maltese Islands, tafoni are dissolution phenomena initiated under a drift mantle or otherwise below ground, subsequently exhumed, and then enlarged by subaerial weathering processes.

INTRODUCTION

This paper describes shallow cave-like features found during a geomorphological survey in the Maltese Islands and assesses a number of potential explanations for them. The features have a consistent morphology and do not seem to result from normal karstic phreatic or vadose dissolution processes, although they are confined to limestone outcrops that in other parts show evidence of karstification, including rillenkarren, true caves and more or less well-developed limestone pavements. The features conform to published descriptions of cavities known as tafoni and cliff foot recesses, which are known from arid regions in a number of countries including Australia (Dragovich, 1967; Bradley et al., 1978), Pakistan (Wilhemy, 1964), Antarctica (Calkin and Cailleux, 1962), Tunisia, Algeria, Morocco, Palestine (Smith, 1978; 1985), the United States of America (Mustoe, 1983), Libya (Peel, 1941; H. Mohamed, Huddersfield University, pers. comm., 1996) and Jordan (author's unpublished data). A similar phenomenon has even been reported from Mars (Smith, 1983).

TERMINOLOGY

Smith (1978) divided shallow bulbous cave-like features which breach the case-hardened faces of cliffs in western North Africa into four classes:

1. 'Cavernously weathered recesses' are bulbous features of the order of 2m high, 3m deep and 4m wide. In the recesses, 'fresh' limestone is exposed. They show many signs of recent weathering, such as flaking, and may have a little comminuted limestone on their floors.

2. 'Seepage recesses' are bulbous features, of similar size to the 'cavernously weathered recesses'. Their roofs and backwalls are commonly covered by deposits of secondary carbonates and iron oxide, and enlarged joints are present in the backwalls.

3. 'Basal tafoni' are small bulbous features, often around 0.5m in diameter, found at the base of cliffs. 'Fresh' limestone is exposed in the backwall, roof and floor of the tafoni.

4. 'Sidewall tafoni' are small bulbous features, again around 0.5m in diameter, found on cliff faces away from the base.

Although there is a clear differentiation between 'seepage recesses' and Smith's (1978) other types, the differentiation between 'basal tafoni' and 'cavernously weathered recesses' is essentially arbitrary and based on the criterion of size only. Likewise, although tafoni are commonly concentrated along the base of cliffs, rapid removal of the talus at the base of a cliffline can lead to former 'basal tafoni' becoming 'sidewall tafoni' (see below).

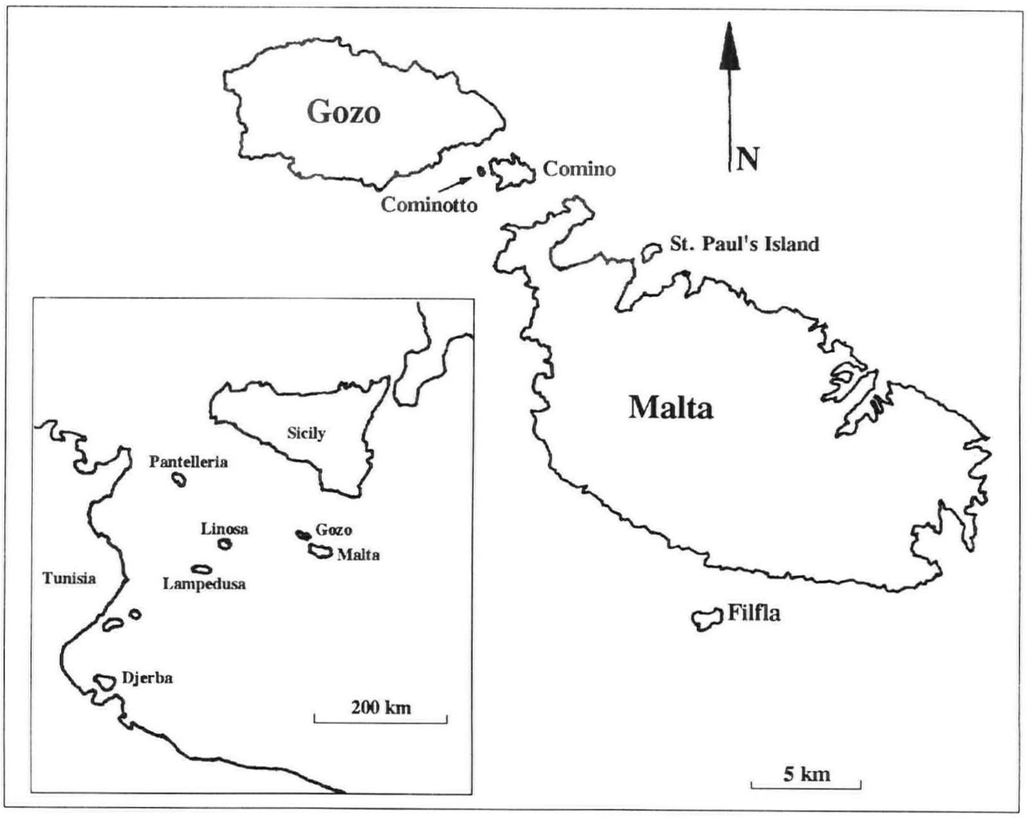


Figure 1. Location of the Maltese Islands. (redrawn from Beckmann, 1987).

Formation	Thickness (m)	Age	Description	Depositional Environment
Upper Coralline Limestone	>160	Messinian-Tortonian	Brown, red and yellow algal limestones with marly partings, oolitic limestones.	Shallow shelf and intertidal, (0-100 m) with algal reefs and lime sand patches, cool subtropical.
Greensand	0-15	Tortonian	Grey-green, weathering orange-brown, sandy limestone, abundant echinoids and giant foraminifera, rare shark teeth and marine mammal bone.	Very shallow shelf (5-40 m), with very slow, episodic deposition, cool subtropical.
Blue Clay	0-70	Serravilian	Blue-grey siltstones and shales, with a few shells.	Deeper (150 m) shelf marine with abundant mud input probably from mountain-building activity to the north.
Globigerina Limestone	22-204	Langhanian-Aquitanian	Buff foraminiferal limestone, massive, with shells and echinoids. 3 phosphatic nodule beds contain corals, shells, echinoids, shark teeth and marine mammal bones.	Deeper (40-200 m) shelf marine, cool subtropical. The nodule beds reflect periods of slow deposition, or even dissolution of the limestone on the sea floor.
Lower Corraline Limestone	>188	Chattian	Pale brown, dense, semicrystalline algal limestone, massive to planar-bedded, locally silicified, with some shell, echinoids and foraminifers.	Shallow shelf (10-200 m) marine, with algal reefs and lime sand patches, cool subtropical.

Table 1. Stratigraphic table for Malta (mostly after Bosence et al., 1981; Pedley et al., 1990).

For the purposes of this paper, therefore, the term 'tafoni' is used to encompass bulbous weathering cavities of all sizes and in all geomorphological positions.

FORMATION OF TAFONI AND SIMILAR FEATURES - A BRIEF REVIEW

A number of processes have been cited to explain the formation of tafoni and basal recesses. It is clear that two major groups of processes lead to these features. 'Seepage recesses' have a different origin from features described by Smith (1978) as 'cavernously weathered recesses' and 'tafoni'. Seepage recesses contain deposits of secondary carbonates and iron oxide. Enlarged joints in the backwall of these features point to solutional processes by 'karst waters'. The deposits of iron oxide and secondary carbonate point to these being relict features that formed in moister conditions than now obtain (Smith, 1978).

'Cavernously weathered recesses', as described by Smith (1978), are currently affected by flaking and granular disintegration, sometimes associated with the presence of salt. Loss of calcium and magnesium from the backwall limestone points to the importance of solutional processes in the initiation of mechanical weathering. These features are suggested to be still forming because their morphology leads to moisture precipitation and absorption, followed by rapid heating and drying, though they were thought to have been initiated in pluvial conditions (Smith, 1978; 1985). A similar origin has been postulated for 'basal tafoni'. Dragovich (1969) and Smith (1978; 1985) hold that 'basal tafoni' are the result of concentrated weathering at ground level, promoted by the availability of dew and frost close to the ground surface. Similar processes were cited to explain 'sidewall tafoni', which are found on cliff faces at places that are structurally weak or particularly susceptible to granular disintegration (Smith, 1978). Once a hollow has formed, this creates a micro-environment in which weathering is concentrated (Smith, 1978; 1985). A similar mechanism, involving the growth of salt crystals in the rock porosity, but not involving solution, was cited by Mustoe (1982) to explain coastal zone 'honeycomb weathering' in the USA.

In western North Africa 'sidewall tafoni' are concentrated in old rockfall scars, perhaps because these were areas of concentrated weathering, or because pressure-release weathering allowed the initiation of tafoni formation, or because the removal of the case-hardening allowed weathering processes to operate (Smith, 1978). Wilhelmy (1964), Winkler (1967) and Nir (in Smith, 1978) also suggest that chemical weathering behind surface case-hardening may initiate tafoni. This process cannot, however, account for the formation of tafoni on rocks on which a surface crust does not form (Bradley et al., 1978). In these situations granular disintegration caused by the growth of salt may be significant.

A further process was suggested by Gilbertson and Hunt (1990), who described small tafoni-like features developed on building stones of a Romano-Libyan farmstead in the Tripolitanian pre-desert. These had clearly formed after the building was erected, and could be demonstrated to have formed just below the ground surface, as sediments accumulated around the walls of the building. They suggested that these weathering features formed as the result of the invasion of the building stones by soil waters containing small concentrations of humic acids during the rainy season, and the subsequent dissolution of intergranular cements by the soil water. Drying of the soil during the summer led to the moisture moving out of the stones by capillarity and evaporating from their surfaces, leading to the reprecipitation of carbonates as a case-hardening. It was clear from the association of these features with the archaeological phasing that this process was extremely rapid, with cavities up to 0.25m across (virtually the size of the stones) forming within 2-3 centuries at most.

THE MALTESE ISLANDS

The Maltese Islands - Malta, Gozo, Comino, Cominotto and Fifla - lie midway between Sicily and Tunisia in the Central Mediterranean (Fig. 1). The islands are predominantly limestone of late Oligocene to Mid Miocene age (Table 1). There is spectacular interstratal karst (Bosence et al., 1981; Pedley, 1990), including a number of circular collapse features up to several hundred metres across. On biogeographical grounds, the islands were probably emergent from the Messinian (Early Miocene) onwards (Giusti, Manganelli and Schembri, 1996; P. J. Schembri, University of Malta, pers. comm., 1995). Structurally, the Maltese Islands consist of a subhorizontal 'layer cake' deformed by a series of dome-like upwarps and basin-like downwarps and cut by a series of NNW-striking faults with vertical displacements of up to 240m and horizontal displacements of several kilometres (Pedley, 1990) into a series of horsts and graben.

Tafoni in Upper Coralline Limestone, Bajda Ridge, Malta.

Physiography in the Maltese Islands is controlled by geological structure and lithotype. Table 2 shows the typical landform assemblages, including karst features, associated with each unit. There are no perennial rivers on the Maltese Islands, though there are a number of short perennial spring-fed watercourses.

The climate of the Maltese Islands is of the usual Mediterranean warm semi-arid type, with predominantly winter rainfall averaging 583mm and with a range of 225-1015mm (Chetcuti et al., 1992). Potential evapotranspiration rates are very high and there is a considerable moisture deficit. Summer maximum temperatures reach 30.6°C. Winter temperatures are around 10°C (Chetcuti et al., 1992; Mayes, 1995). There is strong evidence from Quaternary deposits for periods of severe aridity equivalent to the glacial maxima in northern Europe, and for high rainfall and perennial runoff during interglacials (Hunt, in press).

Formation	Landform assemblage
Upper Coralline Limestone	Steep-sided/cliffed karstified plateaux, limestone pavements, fissures, caves, springs.
Greensand	Steep, rapidly-eroding slopes.
Blue Clay	Steep, rapidly eroding slopes, slumps and shallow slides, chaotic undercliffs on coast.
Globigerina Limestone	Gently undulating tablelands, some large-scale solution features stoped up from L. Coralline Limestone, rare dolines, some water-filled.
Lower Coralline Limestone	High vertical sea cliffs, deeply-incised valleys, often with strongly-developed rillenkarren, limestone pavements, caves, springs.

Table 2. Landform assemblages on the exposed Oligo-Miocene formations in Malta.

METHODS

A number of the features were selected for measurement, with the intention of producing examples from a variety of lithotypes. Actual examples were selected on the grounds of accessibility. Sketch plans were produced by offset survey with the aid of measuring tapes and a plumbline. A number of plans are reproduced in Fig. 2. These indicate features which provide evidence of formational processes. Dimensions of measured features and other characteristics are shown in Table 3.

TAFONI IN THE MALTESE ISLANDS

Tafoni are widespread in the Maltese Islands and can be found on virtually every inland cliff. They can also be seen in the higher parts of some coastal cliffs, especially in softer lithotypes, such as Pleistocene tufa and Globigerina Limestone. Some tafoni, for instance near Manikata and L'Imbordin, are large and stable enough that, with modification, they have been used as troglodyte houses by the Maltese.

The tafoni vary considerably in size (Table 3), from around 0.4m diameter to over 5m in diameter, and there seems to be little pattern to the size distribution. In a number of cases (e.g. 9 and 10 in Table 3), the dimensions of tafoni are controlled vertically by the occurrence of bedding-planes. The bedding planes in these cases seem to be flanked by zones of calcium carbonate induration, which most probably limited the dimensions of the cavity.

In several cases, (1-7, 11), it is possible to demonstrate that the features are developed in surfaces that were relatively recently exhumed. This is so recent in cases 4 and 5 that the fill of these features has been left 'hanging' (Fig. 2). In case 8, where there is no trace of exhumation, the whole feature is case-hardened and thus inactive.

Only a few examples (2-4) had flaky backwalls similar to the 'cavernously weathered recesses' of Smith (1978). Some of these had a fill of angular, unsorted debris, as might be expected with fragments falling from the roof and backwall. The other examples were empty (6, 8), had fine-grained

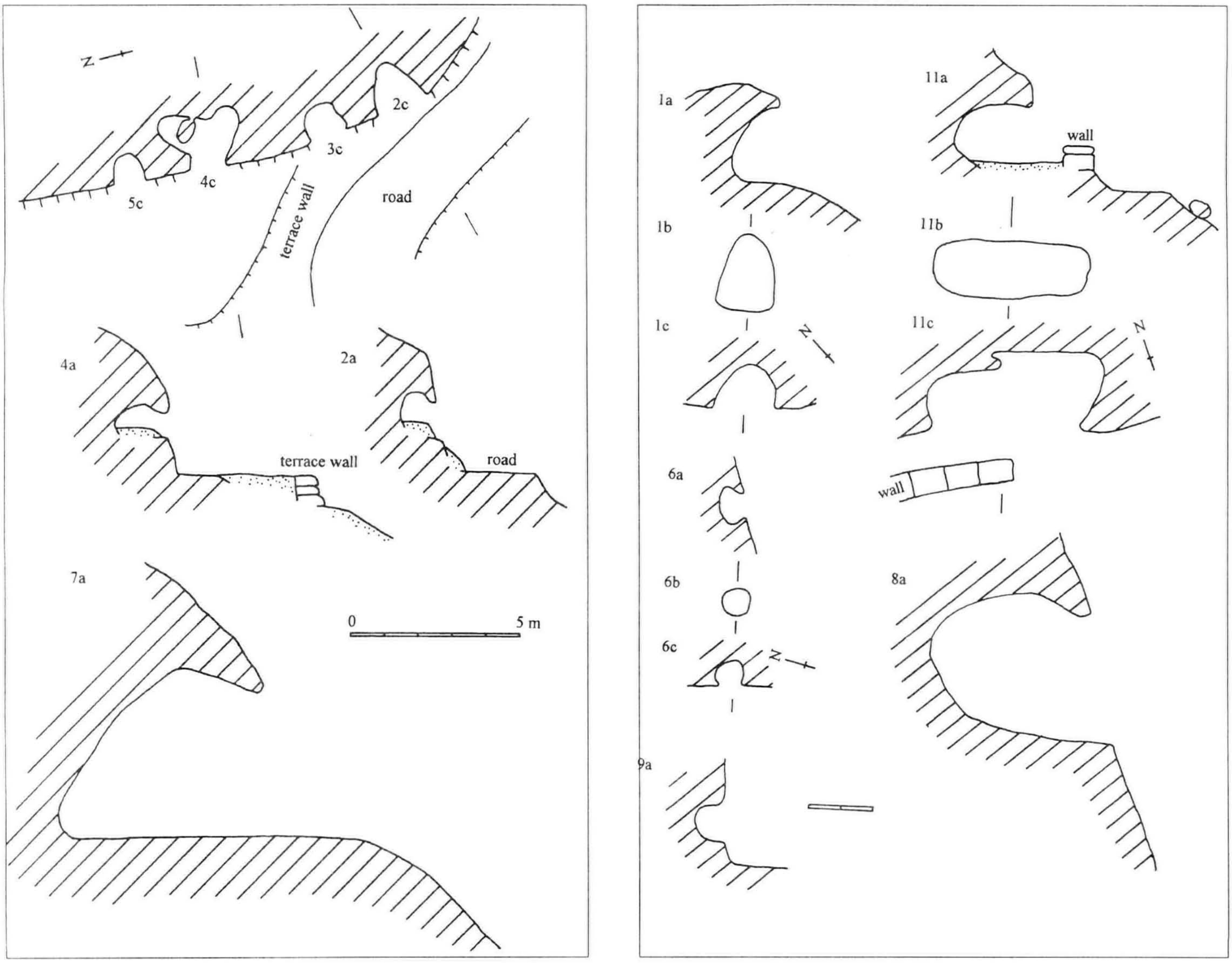


Figure 2. Examples of selected tafoni from the Maltese Islands. The numbers refer to the numbering in Table 3. (a = cross section; b = elevation; c = plan)

fills (7, 11), or were associated with fine-grained superficial materials (9, 10). Several (1, 6, 8-10), are in exposed localities where wind-scour could be expected to contribute to the smoothing of the backwall and the removal of fine-grained debris.

DISCUSSION

As discussed above, the range of sizes of these bulbous weathering features, and the fact that several seem to be relatively recently exhumed, makes the classification of Smith (1978) inappropriate to tafoni in the Maltese Islands. Evidence for the contemporary flaking and granular disintegration of the roof and backwall of several of these features is given by the associated fills (and possibly by associated superficial debris in a number of cases), though one feature was demonstrably relict.

Evidence for the causes of initiation of these features is difficult to locate, but the recent exhumation of several of the tafoni can be demonstrated. This perhaps provides evidence for initiation under the ground. It is suggested that the mechanism proposed by Gilbertson and Hunt (1990) is applicable to these examples, with at least partial dissolution of limestone in the soil water zone followed by exhumation and the rapid enlargement of these features by the mechanisms discussed by Smith (1978), including granular disintegration, salt growth, frost and localised solution. Fine grained debris is most probably removed by aeolian action, which probably also helps to ensure that the morphology of the roof and backwall of the tafoni remain well-rounded. It is entirely possible that a similar mechanism of subsurface weathering followed by exhumation and further development is appropriate for examples elsewhere in the world, for instance in North Africa. Concentration of these features at the foot of slopes would then be consistent with lowering of the ground surface at the foot of cliff lines, for instance by intermittent fluvial activity. Evidence for formerly higher soil/talus lines should be sought to test this proposition.

CONCLUSION

Tafoni are widespread on the Maltese Islands. An assessment of some tafoni leads to the recognition that Smith's (1978) classification of these features is not justified by the Maltese data. It also suggests that a key component of the evolution of these features is partial subsurface solution of the limestone bedrock by soil waters, prior to further subaerial evolution.

ACKNOWLEDGEMENTS

The fieldwork on which this paper is based was done while the author was a visiting scholar in the University of Malta in 1995. I thank Professor P. J. Schembri and Dr. E. Lanfranco for help with the fieldwork and much local knowledge and Professor Schembri and Mr. L. Cassar for arranging the visit to Malta.

Table 3. Dimensions of tafoni and basal recesses in the Maltese Islands.
Dimensions in metres. Grid references are for the Maltese National Grid, except for site 1, which is a protected locality.

	Geomorphological position	Locality	Lithology	Max. height	Max. depth	Max. width	Notes
1	On glacis slope at top of sea cliff.	near M'Taleb, Malta	Pleistocene tufa mound	1.20	0.70	0.95	Back wall very smooth. Partially infilled by angular, unsorted debris. Substrate case-hardened. Exposed by recent soil erosion.
2	In side of road cut.	Kercem, Gozo 303885	Upper Coralline Limestone	1.0	1.0	1.5	Backwall flaky. Partially infilled by angular, unsorted debris. Substrate not case-hardened. Exposed by road cutting.
3	In side of road cut	Kercem, Gozo 303885	Upper Coralline Limestone	0.95	1.0	1.2	Backwall flaky. No fill. Substrate not case-hardened. Exposed by road cutting
4	Above base of small cliff behind agricultural terrace	Kercem Gozo 303885	Upper Coralline Limestone	1.2	1.5	1.2	Backwall flaky. Partially infilled by angular, unsorted debris. Substrate not case-hardened. Exposed by agricultural soil erosion.
5	Above base of small cliff behind agricultural terrace	Kercem Gozo 303885	Upper Coralline Limestone	0.8	0.8	0.9	Backwall flaky. Partially infilled by angular, unsorted debris. Substrate not case-hardened. Exposed by agricultural soil erosion.
6	On slickensided near vertical fault scarp	Kercem, Gozo 302884	Globigerina Limestone	0.4	0.4	0.4	One of about 20 similar features at this locality. Backwall smoothly rounded. No fill. Substrate case-hardened. Exposed by fault movement.
7	Edge of horst, at base of low cliff	L'Imbordin Malta 433739	Upper Coralline Limestone	5.0	7.0	15.0	Backwall smooth. Shallow sandy silt fill. Substrate case-hardened. ?Exposed by soil erosion removing mantle of Quaternary deposits
8	Part-way up small cliff, in gully side	Bajda Ridge, Malta 432782	Upper Coralline Limestone	2.0	2.5	3.0	Backwall smooth and + case-hardened. No fill. Substrate case-hardened.
9	At top of sea cliff, in the side of a mesa-like feature.	Forna Point, Gozo 301934	Globigerina Limestone	0.6	0.4	5.0	One of a number of elongate smoothed features picking out the rock between bedding planes in this locality. No fill. Substrate slightly case hardened. Nearby superficial materials silty sands.
10	In the side of a small mesa-like feature above low sea cliffs	Xwieni Bay, Gozo 324923	Globigerina Limestone	1.0	0.7	3.0	One of a number of similar elongate smoothed features in this locality. No fill. Substrate slightly case-hardened. Nearby superficial materials silty sands.
11	In a road-cut on the glacis slope above a steep coastal cliff	M'Taleb, Malta 414707	Globigerina Limestone	0.9	1.2	1.6	Backwall smooth. A small fill of silty sand. Substrate not visibly case-hardened.

Tafoni in early Holocene tufa, Bajira, Malta

REFERENCES

Bosence, D W J, Pedley, H M and Rose, E P F, 1981. *Field Guide to the mid-Tertiary carbonate facies of the Maltese Islands*. Palaeontological Association, London. 88 pp.

Bradley, W C, Hutton, J T and Twidale, C R, 1978. Role of salts in development of granitic tafoni, South Australia. *Journal of Geology*, 86, 647-654.

Calkin, P and Cailleux, A, 1962. A quantitative study of cavernous weathering (Tafoni) and its application to glacial chronology in Victoria Valley, Antarctica. *Zeitschrift fur Geomorphologie*, N. F. 6, 317-324.

Chetcuti, D, Buhagiar, A, Schembri, P J and Ventura, F, 1992. *The climate of the Maltese Islands: a review.* University of Malta, Malta.108 pp.

Dragovich, D, 1969. The origin of cavernous surfaces (tafoni) in granite rocks of Southern Australia. *Zeitschrift fur Geomorphologie*, N. F. 13, 163-181.

Gilbertson, D D and Hunt, C O, 1990. The UNESCO Libyan Valleys Survey XXI: Geomorphological studies of the Romano-Libyan farm, its floodwater control structures and weathered building stone at site LM4, at the confluence of Wadi el Amud and Wadi Umm el Bagel in the Libyan pre-desert. *Libyan Studies*, 21, 25-42.

Giusti, F, Manganelli, G and Schembri, P J, 1996. *The non-marine molluscs of the Maltese Islands.* Museo Regionale Scienze et Nature Monographie15, Torino.

Hunt, C O, in press. Quaternary Deposits in the Maltese Islands: a Microcosm of Environmental Change in Mediterranean lands. *GeoJournal.*

Mayes, J, 1995. A note on the winter climate of the Mediterranean: a Maltese case study. *Journal of Meteorology*, 20, 323-328.

Mustoe, G E, 1982. The origin of honeycomb weathering. *Geological Society of America Bulletin*, 93, 108-115.

Mustoe, G E, 1983. Cavernous weathering in the Capitol Reef Desert, Utah. *Earth Surface Processes and Landforms*, 8, 517-526.

Pedley, M, 1990. The Tertiary Blue Holes of Gozo *in* Bosence D. (ed.) *Field Guide to the Cenozoic platform carbonates of the Maltese Islands.* 13th International Sedimentological Congress, Nottingham. Field Guide 22, 61-69.

Pedley, M, Bosence, D, Pratt, S and Dart, C, 1990. Tectonic and stratigraphic evolution of the Maltese Islands *in* Bosence D. (ed.) *Field Guide to the Cenozoic platform carbonates of the Maltese Islands.* 13th International Sedimentological Congress, Nottingham. Field Guide 22, 9-39.

Peel, R F, 1941. Denudational landforms in the central Libyan desert. *Journal of Geomorphology*, 4, 3-23.

Smith, B J, 1978. The origin and geomorphic implications of cliff foot recesses and tafoni on limestone hamadas in the northwest Sahara. *Zeitschrift fur Geomorphologie*, 22, 21-43.

Smith, B J, 1985. An integrated approach to the weathering of limestone in an arid area and its role in landscape evolution: a case study from southeast Morocco. *Proceedings, 1st International Geomorphological Conference,* Manchester.

Smith, P J, 1983. Can honeycomb weathering be ET? *Nature*, 301, 291.

Wilhelmy, H, 1964. Cavernous rock surfaces (tafoni) in semi-arid and arid climates. *Pakistan Geographical Review*, 19, 9-13.

CAVE AND KARST SCIENCE Vol. 23, No.2, October 1996
Transactions of the British Cave Research Association

Redefining the boundary between karst and pseudokarst

Charles A. SELF and Graham J. MULLAN
University of Bristol Speleological Society, Students Union, Queens Road, Bristol BS8 1LN

Abstract: The terms "karst" and "pseudokarst" suffer from a plethora of definitions, many of which are mutually contradictory. This paper suggests a non-genetic classification, whereby karst terrain is described by a suite of physical features, without reference to rock type or dominant evolutionary process. Karst-like features in rocks normally considered insoluble would therefore be classed as karst, and this definition could logically be extended to include similar features in superficial deposits. Pseudokarst would then be used to describe landforms such as gulls and lava tubes that have their own unique characteristics and only mimic karst superficially. The sandstone karst of Northumberland and the limestone pseudokarst of the Cotswolds are used as examples.

INTRODUCTION

Widely divergent ideas as to what should be classified as karst have led to a misunderstanding both of that term and the meaning of the term pseudokarst. Prominent among recent examples is a paper by Marker and Swart (1995) in which the cavernous quartzitic sandstones of the Western Cape region of South Africa are classified as pseudokarst on lithological rather than morphological grounds. They also offer genetic definitions of karst and pseudokarst (in this case meaning karst-like features in arenaceous rocks) that give such prominence to dissolutional processes that the two seem essentially the same, although that did not appear to be their intention.

In contrast, Younger and Stunell (1995) used examples of cavernous sandstones and peat piping in Northumberland to support the argument that there is no fundamental difference between pseudokarst and karst. Their use of pseudokarst as a synonym for piping/suffosion caves ignores a number of other pseudokarst landforms. This restricted view is not shared by the caving community and international pseudokarst symposia are dominated by studies of gull caves. The intention of the present paper is to develop the argument of Younger and Stunell and to clarify what may then be meant by pseudokarst.

Before attempting to define pseudokarst it is necessary to review the meaning of the term "karst". This review is intentionally restricted to text book definitions, as previous attempts to clarify the pseudokarst problem (eg. Otvos, 1976) have not been successful. Previous writers have laid stress on various aspects of the karst landscape in their definitions, but have concentrated on two main points: underground drainage and soluble rocks. Thus:

Sweeting (1972) states that the essential characteristic of karst landforms is the presence of vertical and underground drainage and maintains that all karst regions are areas of massive limestones.

White (1988) stresses that karst landscapes and their underlying caves are created by the chemical solution of the bedrock.

Bögli (1980) states explicitly that evaporites and carbonates are karstifiable and allows quartzite karstification only under conditions of extreme tropical humidity and long timescales (silica being sparingly soluble in water).

Jennings (1985) is far more thoughtful and comments that solution is not always the dominant process in karst areas. He discusses karst and pseudokarst in some detail, but still maintains that differences in process in the formation of apparently identical features in different rock types justifies separate classification. Jennings then offers glacier karst as an example of pseudokarst, neatly (if inadvertently) demonstrating the confusion that exists between the two terms.

Jakucs (1977) admits *"...the definition of the karst concept as a state of evolution, dependant on certain conditions, of a mountain-sized mass of limestone does not preclude the manifestation of karst phenomena on other rocks similar in some respect to limestone, under similar influences and conditions"*.

Only Ford and Williams (1989) give a definition independent of process, saying *"...we now consider karst to comprise terrain typically characterised by sinking streams, caves, enclosed depressions, fluted rock outcrops and large springs"*.

In contrast, Parker and Higgins (1990) give a purely process-related definition: karst is the removal of solid soluble rock in solution, pseudokarst is the removal of solid clastic rock in suspension.

This last view is clearly wrong and indeed has been recognised as so for some time. Writing in 1971, Newson stressed the importance of physical erosion in the formation of caves. In the limestone karst areas of the British Isles he found that under base-flow conditions the necessary velocity for abrasion is never reached in swallet fed caves. Also, much of the dissolved load derives from percolation water joining the streamway. With an increase in discharge a different regime is initiated where allochthonous material enters with the swallet water and calcareous sediment derived from the cave walls is discharged from the resurgence. Greater flow rates disproportionally raise the particle size and quantity of sediment that can be transported, with physical erosion increasing by orders of magnitude. The enlargement of some swallet fed stream passages may be primarily due to flood events, though episodic physical erosion is particularly difficult to quantify. A genetic (solutional) definition of karst would place some of the largest limestone river caves of the world in pseudokarst - an absurd notion.

Small-scale features in cave passages may also be produced by physical processes. In the Hölloch (Switzerland), scallops and flutes are concentrated on the floor and lower parts of the cave walls. This shows gravitational control and so must be due to corrasion by water containing suspended grains of sand (Bögli, 1981). If scallops were purely of dissolutional origin they would be more uniformly distributed over the walls and roofs of cave passages. Scallops also form in the beds of surface rivers, regardless of rock type. This does not deny that dissolution contributes to the development of features in limestone that must be formed by purely physical processes in less soluble rock types.

If identical landforms are created by a combination of processes that vary only in degree of importance, they should be classified together. This has some immediate advantages in classifying tropical quartzite terrains where the high rainfall and long time-scale allows karst formation by both dissolution (Bögli, 1980) and by physical erosion (Martini, 1981). Caves in conglomerate are almost always formed by a combination of the two processes. Such caves are of more than just academic interest - the longest cave in Russia, the maze cave Bolshaya Oreshnaya (length 42km) is developed in a limestone/dolomite conglomerate with an arenaceous matrix (Tsikin, 1990).

Nevertheless, physical erosion continues to be largely ignored in karst studies. This is probably because, as Ford and Williams (1989) say, chemical erosion is easier to quantify than physical erosion and therefore more attractive to researchers. In addition, the water industry is primarily interested in the chemical composition of its underground water sources. Thus, lulled by the abundance of solution studies, the majority of karst workers have adopted a view of karst that, by being concerned with one type of process to the exclusion of others, is not only unnecessarily limiting, but also wrong.

In other branches of the Earth Sciences there has been a recognition in recent years that genetic classifications can actually hinder the development of our understanding of processes. Descriptive terminology, on the other hand, will remain unchanged regardless of subsequent changes in the theoretical understanding of processes. We should accept the Ford and Williams definition of a karst terrain with the morphological elements they list - these elements being known individually as karst features and collectively as the karst suite. A karst terrain is therefore a landscape whose dominant features are karst. This does not mean that all elements of the karst suite must be present, but that such features should be numerous. Karst can occur as occasional features in other terrains, due to local factors, and may also be found deep below the surface (palaeokarst, hydrothermal karst), where it is independent of the surface terrain. Karst is, therefore, used as an adjective to describe certain morphological elements, and as a noun to describe the landscape itself. "Karstic" is an alternative adjectival form in common usage.

In the following two sections the morphological features of two contrasting terrains in the British Isles are examined (Fig. 1). These demonstrate that all the features of the karst suite can be formed in a short time in a virtually insoluble rock type, whilst superficially similar features, even when found in limestone, do not necessarily indicate a karst landscape.

NORTHUMBERLAND: KARST WITHOUT CARBONATE

A major spur to the writing of this study was a paper by Younger and Stunell (1995), which used as a study area the Fell Sandstone outcrop of Northumberland. This region has been explored extensively by one of the present authors, (Mullan, 1985, 1989; Mullan and Wilson, 1987; Wilson, 1983, 1991).

The Fell Sandstone is a Viséan (Lower Carboniferous) formation that crops out in an arc stretching from south-east Scotland through Northumberland into the Brampton district of Cumbria. In the study area the formation is between 240m and 300m thick and has a particularly uniform lithology. It is mostly made up of well-sorted, friable, fine- to medium-grained quartzitic sandstone. Cross-bedding is widespread. The sandstone units are commonly separated by thin red and green micro-cross-bedded argillaceous siltstones, with sporadic thin conglomerate bands containing pebbles up to 25mm in diameter. In the context of the current study it must be stressed that it has no significant calcareous content. The outcrop forms barren, heathery moorland hills in a distinctive craggy escarpment around the Cheviot Hills, with a gentle dip slope towards the east, south-east and south (Hodgson and Gardiner, 1971; Robson, 1956). Because of the dip, the outcrop is generally only a few hundred metres wide.

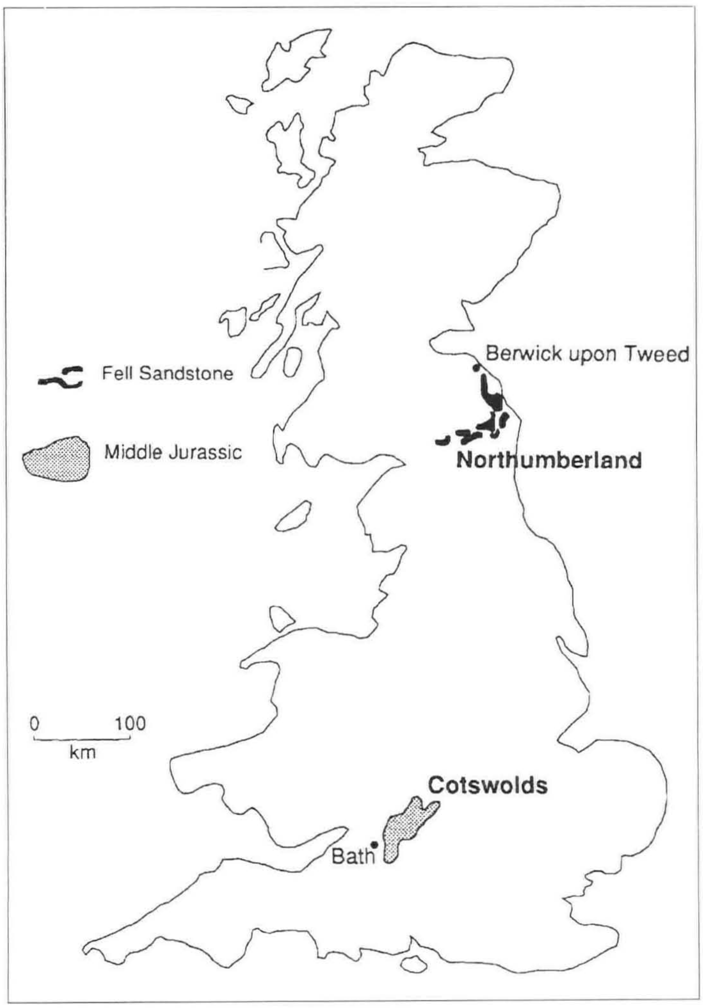
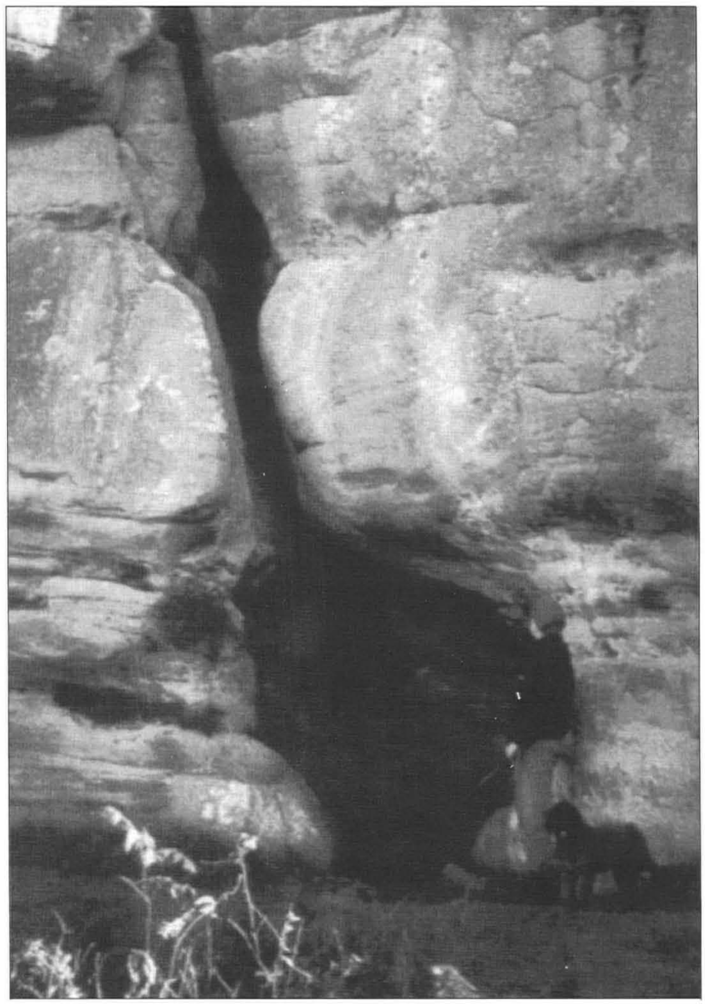


Figure 1. Outcrop of the Fell Sandstone in Northumberland and the Middle Jurassic of the Cotswolds.



Caller Crag cave, Northumberland.

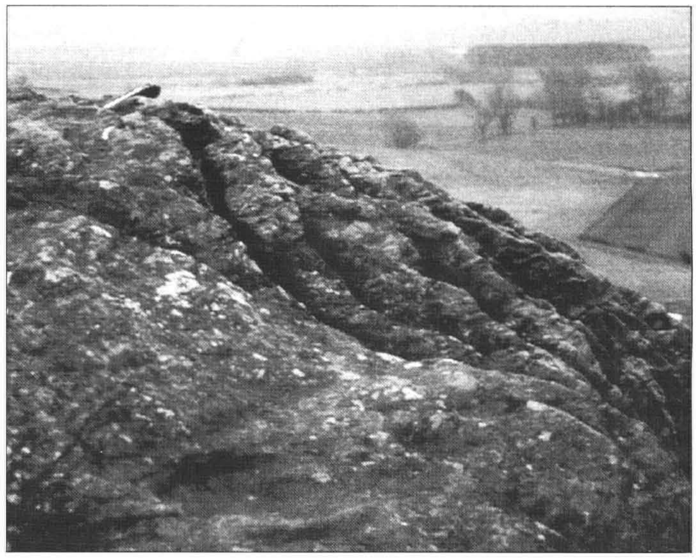
This formation originally came to one author's notice because of its caves but, as exploration and field work progressed, all the other distinctive features of the karst suite were found upon it. A comprehensive investigation of the entire outcrop has not been possible, but the majority of the rock exposures between Berwick on Tweed and Keilder have been visited. Exposures occur at frequent intervals along the length of the outcrop and none has been found that does not demonstrate at least one type of karst feature. Examples of these will be discussed in turn.

Caves

To date more than ten caves have been discovered in this formation, along with a larger number of impenetrably small elliptical tubes. Probably the most distinctive from the point of view of this study is Routin Lynn Cave (NGR NT 982368). This is an elliptical tube developed along a prominent bedding plane and shows distinct signs of formation by flowing water, in that scalloping can be seen. It is, however, only 9.6m long (Mullan, 1989), not the 15m quoted by Younger and Stunell (1995). Most of the other known caves in the area are shorter and wider than this and are to be found in the lines of long crags that are a distinctive feature of the outcrop. Probably the most typical of these is Caller Crag Cave (NGR NU 115070). This cave has developed at the intersection of a south-west/north-east joint and a north-west/south-east joint, giving a chamber with three short continuations at floor level and an upper one leading out onto the top of the crag (Wilson, 1991). The well-known St. Cuthbert's Cave (NGR NU 059353) may be an unusually wide example of this type.

The most unusual find has been Murton Crag Cave (NGR NT 963496), which is a 1m by 0.75m triangular passage that runs for about 10m parallel to the cliff face, between two entrances, though it is blocked by a choke about half way through. It is clearly the product of water action and there is no sign of any mass-movement contribution to its formation. In contrast, Cateran Hole (NGR NU 102236), the longest cave in the Fell Sandstone, is a single rift passage 37.8m long, entirely of mass movement origin (Mullan, 1989).

Small cave entrance, Berryhill Crag.



Rundkarren, Dod Law.

With the exception of the landslip rifts, all the caves have been formed by water action, probably by mechanical erosion (corrasion) following an initial period of chemical weathering. A suggested mechanism is pyrite oxidation in the less pure stratigraphical levels, which initially loosens the sand grains from their matrix, making them susceptible to fluvial transport (Younger and Stunell, 1995).

The primary pathways for water movement must already exist in karstifiable arenaceous rocks as they do in limestone. Lowe (1993) has described how inception horizons develop throughout bodies of limestone over periods of millions, or perhaps tens of millions, of years. These initial pathways can then be exploited by karst groundwater when conditions permit. There is no reason to suppose that such a mechanism should be restricted to carbonates.

Fluted rock outcrops - karren

These features have been noted at several localities within the area. Mullan (1989) found "spongework" and rundkarren at Dod Law (NGR NU 004310) and rundkarren at Routin Lynn; Younger and Stunell (1995) note rillenkarren at St. Cuthberts Cave and on the Duddo Stones (NGR NU 931437), a late Neolithic stone circle made of local sandstone. Later fieldwork has noted similar features at more than eighteen localities in the area. Local tradition has suggested that there was an alternative mechanism for karren formation at Dod Law crag. The karren, it is said, were made by the chains where "the devil hanged his granny" (Mees, A, 1953).

Enclosed depressions

At Deadwater Fell (NGR NY 624980) there is a system of at least 15 dolines. Small streams sink in a number of these, to feed a seepage area at the base of a small cliff. Well-developed caves are also present at this location.

Sinking streams, underground drainage and springs

The aquifer potential of the Fell Sandstone has been recognised for some time, as has the fact of conduit flow within it (Hodgson and Gardiner, 1971; Turner et al, 1993.) All the caves noted above are now dry, but an active stream sink is found at South Yardhope Crag (NGR NT 925005),

Rundkarren, Routin Lynn.

Stream Sink, South Yardhope crag (Fell Sandstone).

where a small stream drops down a 1.5m waterfall in an enlarged joint, to resurge at the bottom of the crag. A map showing the approximate position of the active springs in the study area can be found in Younger and Stunell's (1995) paper.

It can be seen that this virtually carbonate-free landscape contains all the features of the karst suite. However, it does not come into the quartzite karst special case defined by Bögli (1980), as the extremely long timescale (in the millions of years) that he postulated for quartzite dissolution in South America is certainly not applicable here. On the contrary, Younger and Stunell (1995) indicate a timescale for karren formation of only a few thousand years at most. Deeply eroded rillenkarren on the Duddo stones must post-date their erection, some 3,800 years ago. Indeed the region was subject to glacial erosion during the Devensian, which suggests that all the karren forms, and the doline/seepage system on Deadwater Fell, must be strictly post-glacial.

The features forming this suite cannot owe their existence here to the effects of chemical weathering, save in terms of small-scale initiation processes. Their formation has come about through the action of mechanical processes. Despite this, they are morphologically indistinguishable from the same types of feature found on carbonate rocks, and they are indisputably karst. Similar features, notably karren, can be found on a number of other rock types. Such elements may be classified as karst features, but this does not necessarily make their location a karst terrain. It is the presence of all the elements of the karst suite in substantial quantities that allows the Fell Sandstone to be classified in this way.

Resurgence at base of South Yardhope Crag.

THE COTSWOLDS: LIMESTONE BUT NO KARST TERRAIN

The Cotswolds form an elevated tract of land that extends in a south-westerly direction from Chipping Campden in the South Midlands to south of Bath, where they lose their identity as they merge with the dissected upland of the eastern Mendip Hills.

Geologically the Cotswolds comprise an interbedded sequence of mid Jurassic limestones and clays, with some local sandy facies. There is great lateral variation of the beds both in terms of thickness and lithology (Kellaway and Welch, 1948). The two main limestone units, the Inferior Oolite and the Great Oolite, dominate the topography because of their more resistant nature. The gentle south-easterly inclination of the strata gives an extensive dip slope (on which rise the headwaters of the River Thames) and an impressive scarp face overlooking the River Severn.

The Cotswolds are underlain by fine-grained sediments of the upper Lias, which crop out to form low ground to the west of the scarp. Much simplified, the upward geological succession is as follows: Upper Lias (sand and clay), Inferior Oolite (limestone), Fuller's Earth (clay), Great Oolite (limestone), Forest Marble (clay with some limestone units). The dip slope is predominantly of Great Oolite, passing into Forest Marble in the lower ground to the east.

Near Cheltenham, in the north of the region, the Inferior Oolite attains its greatest thickness (90m) and the hills exceed 300m in altitude. Here the Inferior Oolite forms the Cotswold scarp face. In the south of the region, near Bath, a much attenuated Inferior Oolite forms only a minor feature in front of the Great Oolite scarp. These essentially oolitic limestone sequences include massive current-bedded freestones and shelly marine limestones that have been worked extensively for building stone since Roman times. In the southern Cotswolds, the Great Oolite is 30-40m thick and includes the famous Bath Stone.

The surface drainage of the region follows a simple pattern. Rivers rising to the west of the Cotswolds flow to the River Severn. Rivers rising on the dip slope flow to the River Thames. An exception is the (Bristol) River Avon, a scarp stream that has cut through into dip territory, progressively capturing a series of former Thames headstreams (Self, 1995). As a result, the Avon valley has become overdeepened in its middle part, where it cuts through the escarpment.

The combination of permeable and competent limestones overlying impermeable and incompetent strata has led to foundering in the regions of greatest topographic gradient. In many cases the limestone has cambered in the direction of slope, with consequent lateral extension of the strata. Major joints within the rock sequence have opened to produce gulls. This movement does not necessarily affect the near-surface beds, in which case gull caves are formed within the strata. The steepest gradients are found at the scarp edge and on the valley sides above the River Avon and its tributaries.

Hydrology

Both the Inferior Oolite and the Great Oolite are essentially massive limestone units underlain and overlain by impermeable beds. Because of lithological variations within the strata, some of the adjacent beds are locally in hydraulic continuity with the oolites. For this reason, some of the oolite aquifers are a little thicker than expected, but they remain confined by thick clay sequences.

Differences in relief between hilltops and valleys commonly exceed 100m, so a profusion of karst features might be expected. This is not the case. The open joint network of the oolite beds allows rapid groundwater dispersal, commonly to more than one spring, so although drainage is predominantly underground, it is by fissure flow and cave conduits have not developed. Dye tracing experiments give rapid peak travel times (1-5km/day) but extremely long, low-concentration "tails" that continue for up to a month (Smart, 1977). This indicates that the major joints provide conduit flow, but that there is considerable dispersal into subsidiary fissures. No cave passages can be accessed at the springs.

The Cotswold dip slope is deeply dissected by valleys that commonly begin as dry valleys above their perennial springs. These dry valleys are steep sided and many are floored with stony clay of Pleistocene origin. During wet weather they are active due to a combination of run-off from impermeable surface beds and overflow from the unconfined Great Oolite aquifer.

The dip slope also has a number of sinkholes, particularly in areas where a thin covering of Forest Marble masks the Great Oolite beneath. These sinkholes have small catchment areas, mostly field drainage, though some are fed by seepages from limestone beds within the Forest Marble. Pleistocene drift deposits can also provide small catchment areas. Few of these sinks have been investigated by cavers and many have been infilled by farmers. A notable exception is Boxer Pot (NGR SP 115359), where a determined effort enabled a depth of about 40m to be reached in a narrow fissure with no passable continuation at the bottom (Anon, 1969).

The Jurassic limestones of the Cotswolds are important aquifers, supplying drinking water from both springs and boreholes. Though the Cotswolds have subterranean water flow, there is no integrated drainage pattern and therefore no cave streamways.

Gull caves

The combination of competent and strongly jointed limestones overlying soft and impermeable clays has resulted in considerable foundering of the strata in the Cotswold region. The areas most affected are the scarp edge of the Mid and North Cotswolds, and the valley sides of the River Avon and its tributaries. The Geological Survey map shows 45km^2 of disturbed strata in the Bath region.

The foundering of the strata has been by both landslipping and cambering. The latter is of particular interest to cavers, as extension of the strata opens up the joints to produce gulls. Since this mass movement does not extend uniformly through the rock sequence, the near-surface beds commonly remain relatively undisturbed. Gull caves are therefore formed by bed-over-bed sliding within the strata. (For details on the formation of gull caves, see Cooper, 1983; Self, 1986, 1995).

The most extensive gull cave in the Cotswolds lies high on the flank of the River Avon, 4km east of Bath. Sally's Rift (NGR ST 794650) is a complex network of rift passages totalling 350m in length, formed along two jointing directions. This makes it the second longest gull network in Britain (after Buckland's Windypit in North Yorkshire, length 366m).

Sally's Rift is unusual among gull caves in that it has had a geomorphological survey (Self, 1995). Calcite flowstone from the cave has been dated radiometrically, giving ages from 78ka to greater than 350ka. The probable age of the first phase of mass movement is at the end of the Anglian cold stage, when permafrost melting reduced the shear strength of the underlying clays. The cave shows evidence of several phases of mass movement. The cave walls have dissolutional etching of sedimentary features in one of the joint sets, evidence for ancient groundwater movement before the present topography was established (ie. before the development of the valley of the River Avon). Such highlighting of the cross-bedding is more usually seen on surface rocks exposed to wind and rain, where physical processes dominate.

Other significant slip rifts include Isaac's Cave (NGR SO 985256) on Cleeve Hill near Cheltenham (North Cotswold) and Coaley Rift Cave (NGR SO 787997) near Uley (Mid Cotswold). Both are scarp edge caves. Isaac's Cave is a network of perhaps 40m extent (Kay, 1974); Coaley Rift Cave is a single large fissure 80m long (Wintle, 1970). Smaller caves are periodically noted in the journals of local caving clubs.

Despite the huge extent of foundering of the scarp edge and of the valley sides, relatively few major gull caves are known. One reason for this is that gulls are formed within the rock sequence and may have no connection either with the surface above or with the hillslope. Many gulls are only a few metres long, ending abruptly at narrow cross joints. Only when the direction of movement is oblique to the jointing do gull networks form. Gulls close to the surface on steep hillsides tend to be unstable and are commonly filled with bouldery rubble and with surface-derived material. Such gulls remain hidden unless exposed by construction projects or by quarrying, which can include underground stone working.

The Cotswolds is a limestone landscape with underground drainage, but currently this drainage is by unintegrated fissure flow and not through caves. Occasional karst sinks feed the aquifer, but most rainwater enters the ground directly. The only important caves of the region are the pseudokarst gulls, but collapse and infill by material of surface origin prevents access to all but a few of them. Although pseudokarst caves are an important feature of the region, they are not common enough to define even small areas of the Cotswolds as pseudokarst terrain.

The Cotswolds also fail to qualify as a karst terrain, though karst features are undoubtedly present. Lowe and Waltham (1995) use the Cotswolds as an example of fluviokarst in their definition of that term. However, the surface valleys that characterise fluviokarst are the result of low initial permeability before caves develop, or reduced permeability due to ground freezing in a periglacial environment. In both cases, the valleys become dry as karst drainage is established. Smart (1977) has shown that the Cotswold limestones are highly permeable and that the dry valleys are periodically active with overflow from the aquifer. Using Lowe and Waltham's definition of fluviokarst as a type of karst terrain, the Cotswolds seem unsuitable as an example. Perhaps the definition should be extended to include morphologically similar regions (such as the Cotswolds and the Chalk downlands) where high initial permeability hinders cave development. Using non-genetic language, fluviokarst could be separated from karst and described independently as a landscape dominated by surface valleys where a significant number of karst features are also present. In regions where karst drainage is firmly established, fluviokarst would also be a karst terrain. In regions such as the Cotswolds, it would not.

DISCUSSION

Expansion in the meaning of "karst"

In its original meaning, the term karst described the physical features of the Kras region in northern Yugoslavia (mostly within what is now southern Slovenia). The region has some spectacular river caves, but the defining feature of this original "classical karst" is the presence of dolines rather than true valleys. "Kras" is in origin a Slovene word meaning a dry, stony area and "dolina" a general Slav term for a small valley (Šušteršič, 1994). The early usage of karst is described more fully by Kranjc (1994), and the reader should refer to this paper for additional detail.

The term karst was later used to describe other limestone regions outside the old Yugoslavia, though not all had quite the same features. In particular, mountain massifs made entirely of carbonates have no impermeable strata to act as a rainfall collector, and so have no sinking streams. Great differences in altitude between rainwater infiltration points on the mountain tops and risings in the valleys creates a different, essentially vertical, style of cave. High on the mountain the surface rocks may be frost-shattered, with stone-choked potholes the only obvious karst features. Such a terrain is commonly called "alpine karst".

Karst has proved a very versatile term and differences between karst terrains have had to be distinguished by prefixes. "Cone karst" is a landscape of low conical hills formed in wet tropical climates, with broken valleys or dolines between them draining into sinkholes. "Tower karst", which is such a spectacular feature of Guangxi province in China, forms from residual cones steepened by water table undercutting from surrounding alluvial plains (Lowe and Waltham, 1995).

Fossil landscapes have also been brought within karst, with "palaeokarst" used to describe features that were formed during past geological periods. "Buried karst" is used to define karst features of surface origin, buried under later sediments, and "intrastratal karst" for solution features formed at depth (Bosak, Ford and Glazek, 1989). Such deep circulating water may be of meteoric, petroleum or connate origin and can have a considerable dissolved load and/or thermal energy (hydrothermal karst). Intrastratal caves formed by hot brines are commonly filled by mineral deposits, particularly sulphides (eg. Mississippi Valley Type lead and zinc deposits).

Morphologically, phreatic caves formed by thermal and by cold meteoric water have much in common. Similarities can also be found in the vadose zone. At shallow depth, thermal water caves can develop upwards from the water table by condensation corrosion (Szunyogh, 1990), a process that is seen to a lesser extent in limestone caves formed by meteoric water. In more soluble rocks, condensation corrosion can be the main cave-forming process; in the gypsum karst flanking the Kugitangtau hills of eastern Turkmenia, warm groundwater and very cold winter air temperatures produce enough convection for rapid upward cave growth (Maltsev and Self, 1992). In Boy Bulak limestone cave in Uzbekistan (depth 1415m), condensation water in the topmost part of the system collects together to form the small stream seen at the entrance (Bernabei and De Vivo, 1992). This is still a karst stream, even though the water was generated within the cave.

Condensation corrosion becomes much more pronounced if there are acidic gases present in the cave air, the most important of which are sulphur dioxide and hydrogen sulphide. These gases are commonly present in water of hydrothermal origin, and are also produced by sulphur bacteria in many different types of cave. Condensation and bacterial processes must also be present in caves formed in arenaceous rocks, but their effect is very small. Theoretically they could assist in the chemical breakdown of the bedrock cement, but the main cave-forming agents are the traditional ones of sinking streams and rainwater infiltration.

The meaning of karst has also been extended beyond limestone to describe terrains in dolomite and other carbonate rocks. Similar features in soluble but non-carbonate rocks are usually described by putting the mineral name as a prefix, eg. "gypsum karst". These more soluble minerals form only a limited range of karst features compared to carbonates, and are unfavourable for the preservation of palaeokarst. The softness and plasticity under pressure of evaporite minerals permits rapid destruction of any karst forms. An illustration of this is the Iron Age salt mine of Hallstatt in Austria, where the old adits have disappeared by annealing, leaving the mining tools buried within the salt (Bögli, 1980). Gypsum and anhydrite landforms are preserved only rarely in the palaeokarst record. Chemical transformation between the two minerals, with accompanying volume changes, occurs during both burial and uplift (Bosak, 1989). Palaeokarst features are preserved mostly in carbonates and in arenaceous rocks.

When describing collapse structures of palaeokarst origin in Canada, Ford (1989) declared: *"...the Elk Point salt would appear to be one of the world's great interstratal karsts"*. This statement is very far removed from the original description of karst as a suite of landforms in a part of Slovenia. Karst as a term has evolved and expanded far beyond its origins, encompassing most forms of solutional erosion. Yet the study of the physical erosion aspect of karst has hardly evolved at all. When karst caves have been found in quartzitic rocks, long-timescale dissolutional origin is often presumed (Bögli, 1980; Marker and Swart, 1995), even though Martini (1981) has shown that, in the Transvaal region of South Africa, such processes merely loosen the intergranular cements, and physical erosion is responsible for cave formation.

Karst landforms are created by both chemical and physical processes. Newson (1971) has shown that, for limestone, the balance between dissolution and physical erosion changes as a karst conduit develops, though the two processes produce essentially the same features. Karst landforms in evaporites and in arenaceous rocks are simply extreme members of a continuum of such features. If hydrothermal mineral deposits in limestone can be accepted into karst, surely lithology alone should not be allowed to prevent the course of a river that sinks and flows through a quartzite massif being described as a karst streamway.

If karst-like features in arenaceous rocks are to be accepted into karst, regardless of their genetic origin, a case can also be made for inclusion of subterranean conduits in superficial deposits. Younger and Stunell (1995) have already suggested this, noting that pipe features in soil have a similar structure and hydrological function to caves in consolidated rocks. In the particular case of pipes through peat, the erosion products tend to be transported as humic colloids rather than as suspended particles. The comparison with caves of dissolutional origin is hard to avoid. There is, however, an interesting contrast between the long timescales required for cave formation in compact quartzitic sandstones and the very short survival times, to a matter of months, reported for some caves formed by piping in poorly consolidated sandstone (Pearman, 1989).

Glaciers, particularly in the high Arctic, can produce an even more karst-like terrain, with meltwater streams sinking into glaciers and cave and pothole systems forming within the ice. Rivers in meandering passages with scalloped ice walls flow at the base of the glacier, emerging at the snout as major springs. Extensive cave systems explored within the Werenskiold and Torrell glaciers of Spitzbergen (Rehak and Ouhrabka, 1989) have many of the features of classic karst caves, providing the "rock type" is ignored.

In general, the movement of water through channels in a solid material tends to produce features that are recognisably karstic. These features are the result of turbulence in the stream flow, so similar features are produced whatever the mechanism of removal of material from the walls, whether corrosion, corrasion or even melting. A karst definition independent of process would allow field descriptions to be made without the risk of error and future re-interpretation. Karst landforms should not be "tied" to any one rock type or any one process, and the common habit of regarding any outcrop of limestone as karst is wrong.

Karst is thus the general term used to describe certain physical features in the surface and near-surface environments; namely sinking streams, caves, enclosed depressions, fluted rock outcrops and large springs.

Collectively these features are known as the karst suite. A landscape can be called a karst terrain if the dominant elements are karst. Karst is found in many different rock types and can be formed by different processes. The term karst is also used to describe deep-seated dissolution features.

Pseudokarst

What then of pseudokarst? There is still a need for such a term to describe features that are superficially similar to karst, but have some unique element that is not found in karst. Tectonic caves are the result of extension of the strata, unlike karst caves, which are produced by erosion; the cave passages have characteristic asymmetry (fit features) that cannot be formed in erosional conduits. Gull caves are produced by extension within the strata during cambering and have similar asymmetric features to tectonic caves; vertical displacement between the walls is also a common feature. Lava tubes form while the host basaltic magma is solidifying, and they are therefore primary features (syngenetic). Some sea caves have a characteristic form due to the surging action of waves. Talus caves have an unambiguous morphology, whilst caves of wind and frost origin, eg. desert rock arches, are similarly distinctive.

Pseudokarst may be regarded as a "broad church" term for superficially karst-like features that can nevertheless be distinguished easily from karst. Such features are not normally found together so, unlike karst, there is no "pseudokarst suite". Pseudokarst is commonly only an occasional element in terrains dominated by other features. Only rarely do such elements describe a terrain themselves, eg. thermokarst: a periglacial landform of enclosed depressions caused by the selective thaw of ground ice.

Karst and pseudokarst are not mutually exclusive; cliff-edge gulls and sea caves are found in karst terrains, and occasional sinking streams are found in otherwise non-karst terrains. On the Isle of Portland (Dorset), ancient karst caves have been dissected by recent gulling to produce polygenetic cave networks (Cooper et al, 1995).

Pseudokarst covers a number of quite different features that superficially resemble karst in some way. Such features have unique elements that clearly separate them from karst. Pseudokarst terrains are known, notably in the periglacial environment, but pseudokarst is normally represented by sporadic features within other terrains.

THE FUTURE

What of the future of the terms "karst" and "pseudokarst"? Perhaps the future is already here, since the NASA Viking Space Mission transmitted photographs of piping, thermokarst and probable lava tubes on the planet Mars (Baker, 1981). Though entitled *"Pseudokarst on Mars"*, Baker's paper identified many features that this review would call karst (eg. valleys that begin and/or end in depressions). In the more distant future the possibility must be considered of karst-like features being found on the outer planets of the solar system, or their moons. Should this most versatile of terms expand yet again, perhaps to include "ammonia karst"?

REFERENCES

Anon., 1969. Boxer Pot, near Broadway. *Gloucester Speleological Society Newsletter*, Aug-Sept, 5-6.

Baker, V R, 1981. Pseudokarst on Mars. *Proceedings of the 8th International Congress of Speleology*, Bowling Green, Kentucky, USA, Vol.1, 63-65.

Bernabei, T and De Vivo, A, 1992. Grotte e Storie dell'Asia Centrale. *Centro Editoriale Veneto*. 307 pp. (In Italian and English).

Bosak, P, 1989. Problems of the origin and fossilization of karst forms. 577-598 *in* Bosak, P, (ed.), *Palaeokarst - systematic and regional review*. Academia Prague and Elsevier, Amsterdam.

Bosak, P, Ford, D C and Glazek, J. 1989. Terminology. 25-32 *in* Bosak, P, (ed.), *Palaeokarst - systematic and regional review*. Academia Prague and Elsevier, Amsterdam.

Bögli, A, 1980. *Karst Hydrology and Physical Speleology*. Springer-Verlag, Berlin. 284 pp.

Bögli, A, 1981. Scallops. *Proceedings of the 8th International Congress of Speleology*, Bowling Green, USA , 1, 82-3.

Cooper, R G, 1983. Mass movement caves in Great Britain. *Studies in Speleology*, Vol. 4, 37-44.

Cooper, R G, Graham, N and Read, M, 1995. Solutional cave passages intersected by mass movement rifts in the Isle of Portland, Dorset. *Studies in Speleology*, Vol.10, 29-35.

Ford, D C, 1989. Palaeokarst of Canada. 313-336 *in* Bosak, P, (ed.), *Palaeokarst - systematic and regional review*. Academia Prague and Elsevier, Amsterdam.

Ford, D C and Williams, P W, 1989. *Karst Geomorphology and Hydrology*. Unwin Hyman, London. 601 pp.

Hodgson, A V and Gardiner, M D, 1971. An investigation of the aquifer potential of the Fell Sandstone of Northumberland. *Quarterly Journal of Engineering Geology*, Vol. 4, 91-109.

Jakucs, L, 1977. *Morphogenetics of Karst Regions*. Adam Hilger, Bristol. 284 pp.

Jennings, J N 1985. *Karst Geomorphology*. Basil Blackwell. Oxford. 293pp.

Kay, R, 1974. Isaac's Cave. *Journal of Sheffield University Speleological Society*, Vol. 2, No.3, 22-25.

Kellaway, G A and Welch, F B A, 1948. *British Regional Geology. Bristol and Gloucester District*. Institute of Geological Sciences, H.M.S.O., London. 91 pp.

Kranjc, A, 1994. About the Name and the History of the Region Kras. *Acta Carsologica*, Vol. 23, 81-89

Lowe, D J, 1993. The Forest of Dean caves and karst: Inception Horizons and iron ore deposits. *Cave Science*, Vol. 20, 31-43.

Lowe, D and Waltham, A, 1995. A Dictionary of Karst and Caves. *British Cave Research Association. Cave Studies Series*, No.6. 41 pp.

Maltsev, V A and Self, C A, 1992. Cupp-Coutunn cave system, Turkmenia, Central Asia. *Proceedings of the University of Bristol SpelFological Society*. Vol.19, 2, 117-149.

Marker, M E and Swart, P G, 1995. Pseudokarst in the Western Cape, South Africa: Its paleoenvironmental significance. *Cave and Karst Science*, Vol.22, No.1, 31-38.

Martini, J, 1981. The control of karst development with reference to the formation of caves in poorly soluble rocks in the Eastern Transvaal , South Africa. *Proceedings of the 8th International Congress of Speleology*, Bowling Green, USA, 1, 4-5.

Mees, A, 1953. *"Northumberland"*. King's England series, 93-94.

Mullan, G J, 1985. The most northerly cave in England. *Newsletter of the University of Bristol Speleological Society*, Vol. 2, 6, 5.

Mullan, G J, 1989. Caves in the Fell Sandstone of Northumberland. *Proceedings of the University of Bristol Speleological Society*, Vol.18, 3, 430-437.

Mullan, G J and Wilson, L J, 1987. The Caves of North Northumberland Volume Three. *Newsletter of the University of Bristol Speleological Society*, Vol.3, 3, 11-12.

Newson, M D, 1971. The role of abrasion in cavern development. *Transactions of the Cave Research Group of Great Britain*, Vol.13, No.2, 101-107.

Otvos, I G, 1976. "Pseudokarst" and "pseudokarst terrains" : Problems of terminology. *Bulletin of the Geological Society of America*, Vol. 87, 1021-1027.

Parker, G G Sr and Higgins, C G, 1990. Piping and pseudokarst in drylands (with case studies by Parker, G G Sr and Wood, W W). 77-110 *in* Higgins, C G and Coates, D R (eds), Groundwater Geomorphology: The role of Subsurface Water in Earth-Surface Processes and Landforms. *Geological Society of America Special Paper 252.*

Pearman, H, 1989. Sand Caves! *Newsletter of the Chelsea Speleological Society*, 32, 1, 2-3.

Rehak, J and Ouhrabka, V, 1989. Spicberky '88. Speleoforum '89. *Ceska Speleologicka Spolecnost*, 21-25. (in Czech with English abstract).

Robson, D A, 1956. Sedimentary study of the Fell Sandstones of the Coquet Valley, Northumberland. *Quarterly Journal of the Geological Society of London*, Vol.112, 241-262.

Self, C A, 1986. Two gull caves from the Wiltshire/Avon border. *Proceedings of the University of Bristol Speleological Society*, Vol.17, 2, 153-174.

Self, C A, 1995. The relationship between the gull cave Sally's Rift and the development of the River Avon east of Bath. *Proceedings of the University of Bristol Speleological Society*, Vol.20, 2, 91-108.

Smart, P L, 1977. Jurassic aquifers of the southern Cotswolds. *Excursion Guide K2: Southern England and South Wales, 7th International Congress of Speleology, Sheffield 1977*, 49-59.

Šušteršič, F, 1994. Classic Dolines of Classical Site. *Acta Carsologica*, 23, 123-156.

Sweeting, M M, 1972. *Karst Landforms*. The MacMillan Press, London. 362 pp.

Szunyogh, G, 1990. Theoretical investigation of the development of spherical niches of thermal water origin - second approximation. *Proceedings of the 10th International Congress of Speleology, Budapest 1989*, 3, 766-768.

Tsikin, R A, 1990. Karst of Siberia. *Krasnoyarsk. Izdatelstvo Krasnoyarskogo Universiteta*. 155 pp. (in Russian).

Turner, B R, Younger, P L and Fordham, C E, 1993. Fell Sandstone Group lithostratigraphy south-west of Berwick-upon-Tweed: Implications for the regional development of the Fell Sandstone. *Proceedings of the Yorkshire Geological Society*, Vol. 49, 269-281.

White, W B, 1988. *Geomorphology and Hydrology of Karst Terrains*. Oxford University Press, New York. 464 pp.

Wintle, J, 1970. Coaley Rift Cave. *Newsletter of the Severn Valley Caving Club*, Oct-Nov, 3-6.

Wilson, L J, 1983. St. Cuthbert's Cave. *Newsletter of the University of Bristol Speleological Society*, Vol. 2, 1, 5-6.

Wilson, L J, 1991. The Caves of North Northumberland Volume Four. *Newsletter of the University of Bristol Speleological Society*, Vol.7, 1, 6-7.

Younger, P L and Stunell, J M, 1995. Karst and Pseudokarst: An Artificial Distinction? 121-142 *in* Brown, A G (ed.), *Geomorphology and Groundwater*. John Wiley and Sons Ltd., London.

CAVE AND KARST SCIENCE Vol. 23, No.2, October 1996
Transactions of the British Cave Research Association

3-D vector processing of magnetometer and inclinometer data

David GIBSON
Department of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT

Abstract: Cave surveying currently relies on visually reading a compass and clinometer. The price of magnetometers and solid-state inclinometers is falling rapidly; thus an electronic compass/clinometer with automatic data-logging is now feasible at an affordable price. However, the interpretation of the readings needs to be done with care because if the instruments are not gimballed then small errors in "pitch" and "roll" can have noticeable affects on the accuracy, with a 1° error in elevation and roll giving rise to a bearing error of over 3° in some cases. Surprisingly, perhaps, more than three axes of measurement are required to compensate for this, and for other sources of error - four axes are essential, and five or six are an advantage. The 3-D processing of this information is straight forward to implement, but difficult to derive. This paper gives formulae for the processing of multi-axis information, and explains the derivation using 3x3 rotation matrices. This paper was first presented at the BCRA Science Symposium in February 1996.

INTRODUCTION

The use of electronic instruments in cave surveying was reviewed in an earlier paper (Gibson, 1996a). This covered techniques such as 'global positioning' by time-of-flight measurements from surface transmitters, and inertial navigation as well as the traditional 'point-to-point' methods. The traditional type of cave surveying, using a compass, clinometer and tape can be made easier by using electronic instruments and a data-logger. Such instruments are expensive, but the basic sensors are now available at an affordable price. Some instruments and sensors were reviewed by Gibson (1996b). The aim of this paper is to describe, to an electronic engineer, the essential signal processing that is required if one is to design a cave-surveying tool around 'raw' sensors.

COMPASS AND CLINOMETER SENSORS

Methods of reading a mechanical compass and clinometer electronically were discussed by Gibson (1996a). Here we will assume that electronic sensors are to be used. In their basic form these will comprise magnetometers to sense the earth's magnetic field and tilt or inclinometer sensors to detect the gravity field.

Magnetometers

A brief description of devices suitable for measuring the earth's magnetic field can be found in a geophysics textbook such as Telford et al (1976). The two devices most suitable to construct are the proton magnetometer and the flux-gate magnetometer.

The proton magnetometer utilises the phenomenon of nuclear magnetic resonance or, more specifically, the precession of proton spin in a magnetic field. The physical principles of the device are outside the scope of this paper; but although they may be unfamiliar to an electronic engineer they are quite simple and such a device should, in theory, be straightforward to construct. A basic proton magnetometer measures only the field magnitude and not its direction. However it is possible to turn it into a 'vector' device by a suitable arrangement of the energising coils.

The flux-gate magnetometer operates by measuring the magnetic field needed to saturate a small piece of ferrite. This depends on the value of the 'background' flux from the earth's field. The flux-gate device is simple in concept, but is notoriously difficult to set up correctly. It is, for example, very prone to temperature drift. The advantage is that it is, inherently, a vector device, and it can be made very small. Recently, a version of the flux-gate has been designed that does not suffer from temperature effects and which is therefore much easier to set up. The device can measure the earth's field to the accuracy and resolution we require (see Gibson, 1996b), and would be a good choice to use as the sensing element of an electronic compass.

Inclinometers

Inclinometers of several types and constructions are common in industry. One version is simply a weighted optical encoder disc. Another type is the so-called 'electrolytic' design that utilises the flow of liquid in a tube fitted with electrodes; measuring its varying resistance. The most compact inclinometers are based around a micro-machined silicon capacitance cell. The devices are available in a standard IC package. They are essentially accelerometers, but in addition to responding to dynamic forces they will measure the 'static' acceleration due to the earth's field. Since they are vector devices they can sense 'tilt', and they are marketed for this purpose.

DESIGNING A 2-D COMPASS

It is a simple operation to use two vector magnetometers at right-angles to measure 'x' and 'y' components of the earth's field. Consider an orientation where both sensors are horizontal and give a maximum positive output when pointed north. The outputs can then be written:

$$\left.\begin{aligned} X &= -A\sin\theta \\ Y &= A\cos\theta \end{aligned}\right\} \qquad (1)$$

where we assume that the sensors are aligned along local x' and y' axes; that the gain A is the same for both, and that the bearing θ is measured clockwise from magnetic north to the y' axis ($x' = 0$) (Fig. 1).

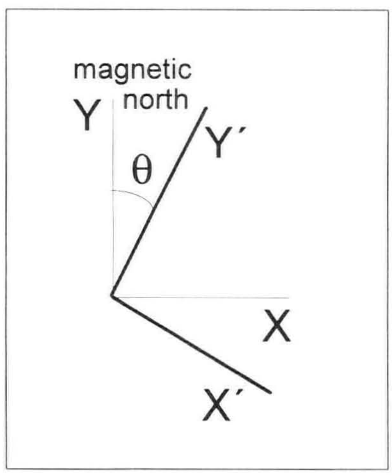


Figure 1. 2-D compass orientation.

The sensor outputs can be processed in software to give the compass heading, from the relationship:

$$\frac{X}{Y} = \frac{-A\sin\theta}{A\cos\theta} \qquad (2)$$
$$\Rightarrow \theta = \tan^{-1}\left(-\frac{X}{Y}\right)$$

The inverse tangent function is easily calculated in a microprocessor using the infinite series:

$$\tan^{-1}x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \ldots,\ \left(|x| < 1\right) \qquad (3)$$

but memory is cheap enough that a look-up table is easy to implement. (A table of 16-bit values of sin$x$ from 0 to 90° in increments of 0.01° would use under 18Kb). Clearly, we need to process the quadrant information (the sign of X and Y) separately; and for angles where $|X/Y| > 1$ we use $\theta = 90° - \tan^{-1}(Y/X)$. We also need to remember that (3) gives θ in radians.

Another method of calculating θ is to make use of a servo algorithm similar to that used in a resolver to digital converter (RDC). A resolver is a standard piece of angular positioning equipment that produces two analogue signal outputs sinθ and cosθ. The RDC "guesses" the angle θ' and produces in hardware the product/sum:

$$A\sin\theta\cos\theta' - A\cos\theta\sin\theta' \qquad (4)$$
$$= A\sin(\theta - \theta')$$

For small errors this produces a signal which is linearly proportional to the error. This is used to drive a feedback loop (a voltage-controlled oscillator, counter and ROM look-up table) which 'servos' to bring the error to zero. The operation is fast and accurate, and does not depend on the value of A. Because it only involves multiplication operations it is ideal for implementation in a digital signal processor. The control loop is second order so it requires compensation, but this is a standard technique covered in the data-sheets. (Semiconductor manufacturer *Analog Devices* produces a range of RDCs, together with Applications Notes explaining their operation). Ten years ago RDCs were manufactured as thin-film hybrid modules and were considerably expensive but, in recent years, cheaper monolithic designs have become available. The signals from the magnetometers might not be in precisely the form that the RDC requires, so it may not be an advantage to specifically make use of an RDC chip; however the same technique can be implemented in software. Even if a full servo algorithm is not used, a 'trial and error' approach may converge on the result quickly enough for our purposes.

ERROR ANALYSIS

If the XY compass moves out of a horizontal plane it will be subject to errors, as we will discuss later. Assuming, for the moment, that it remains horizontal there are three sources of error we can consider:

- conversion error (quantisation)
- gain mismatch
- axis alignment error

Conversion error (quantisation)

Quantisation error is simple to deal with. Suppose the sensor outputs are digitised to 8 bits in a sign/magnitude format (i.e. a count of ±127). Ignoring any rounding algorithms, we can say, to a fair approximation, that the ADC output will be a count of 90 for angles of 45°±0.3°. The resolution, at the 45° point is therefore 0.6°. At 90° the resolution is slightly better. To err on the safe side, and to ease the calibration routines we can use a higher resolution - 12 and 14 bit DACs are not expensive. A 12 bit digitisation would give a resolution of 0.04°. Conversely, and perhaps surprisingly, with only 6 bits we can still achieve a resolution of 2.9° or better.

Gain mismatch

If the two channels have differing gains this will give rise to an angular error that varies with position. Suppose the X channel reads 5% too high, and the Y channel reads 5% too low. The errors are then:

θ	apparent θ	error
0	0	0
45°	47.86°	+2.86
90°	90°	0
135°	132.14°	-2.86
180°	180°	0
225°	227.86°	+2.86
270°	270°	0
315°	312.14°	-2.86°

If the gain can be trimmed to one part in 127, or ±0.8% then the maximum angular error is ±0.6°. With a 12 bit ADC (gain trimmed to 1 in 2047 or 0.05%) we would have ±0.03°.

A similar error arises if there are offsets in the signals recovered from the sensors and, of course, there is the linearity to consider. With good quality sensors the probiems will not be severe. State-of-the-art low-drift, low-offset op-amps, together with software self-calibration algorithms, will add to the accuracy and resolution.

Axis alignment error

If the two magnetometers are not exactly at right angles then each will pick up some for the signal destined for the other. If the angle between the sensors is too small by an amount ε (Fig. 2), then equation 1 becomes:

$$\left.\begin{aligned} X &= -A(\sin\theta\cos\varepsilon - \cos\theta\sin\varepsilon) \\ Y &= A\cos\theta \end{aligned}\right\} \qquad (5)$$

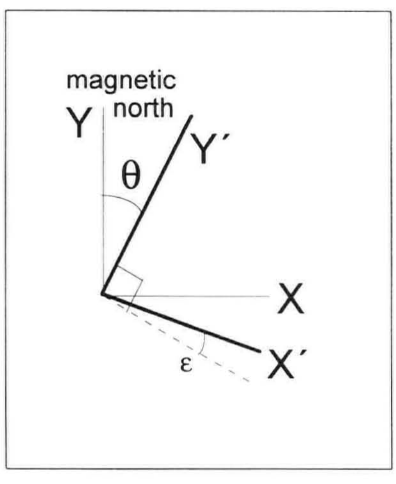


Figure 2. 2-D axis-alignment errors.

So, if ε is measured during a setting-up and calibration routine, we can easily compensate for it. In fact, if $\varepsilon < 5°$ then $\cos\varepsilon \approx 1$ and $\sin\varepsilon \approx \varepsilon$ to within 0.4% (ε in radians) so that:

$$\frac{X}{Y} = \frac{-A(\sin\theta - \varepsilon\cos\theta)}{A\cos\theta} \qquad (6)$$
$$\Rightarrow \theta = \tan^{-1}\left(\varepsilon - \frac{X}{Y}\right)$$

This simplifies the arithmetic a little, but even with the full expression, the problem is hardly insoluble. The treatment of axis misalignment can be extended to the case where both axes are at an unknown angle to the heading indicator, whilst being approximately normal to each other. A calibration routine, written in software, can enable axis alignment to be determined and corrected for. A problem related to alignment and signal offset, and which can also be calibrated out, is the 'orthogonality' of the sensor, by which we mean the situation where the sensor gives a small output even when the field is normal to it.

3-D VECTOR MANIPULATION

We have seen that a 2-D electronic compass is straightforward to design, but it was stated that it needed to be held horizontally. This fact is easily verified. If the Y-magnetometer is pointed to magnetic north, so that the X-axis points east, then rolling the compass east introduces an erroneous X reading whilst the Y reading is unaltered. Similarly, with the compass (Y axis) pointing east, an elevation introduces an erroneous Y reading whilst the X reading remains unaltered.

To quantify the effect we need to obtain an expression for the X, Y and Z fields in terms of the heading and elevation of the compass. There is a third parameter - the 'roll' - which we also have to take into account. We begin by defining the axes - the fixed axes will be X (due east), Y (due north) and a third, Z axis vertically upwards (Fig. 3). The compass's local axes, X', Y' and Z' will coincide when we point the compass due north, and hold it level. The **bearing** or **heading** θ is a rotation clockwise about the Z-axis, viewed from above. The **elevation** or **pitch** ϕ is a tilt upwards, about the *local* X'-axis. The **roll** ψ is a twist clockwise about the *local* Y'-axis. These terms are some of the many that are used to describe 3-D rotations - some other terms are explained in the glossary (Fig. 4). If each of these transformations is applied singly it is straightforward to represent them as operations on the fixed axes. The transformations are commonplace in computer graphics but their derivation may be unfamiliar outside that field. In matrix-vector terms the transformations are:

i) **Bearing/Heading.** Rotate (clockwise viewed from above) about fixed Z axis by angle θ:

$$\mathbf{HEAD} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (7)$$

ii) **Elevation/Pitch.** Rotate ('up' viewed from side) about fixed X-axis by angle ϕ

$$\mathbf{PITCH} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{pmatrix} \qquad (8)$$

iii) **Roll** (clockwise viewed from y<0) about fixed Y-axis by angle ψ

$$\mathbf{ROLL} = \begin{pmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{pmatrix} \qquad (9)$$

Our initial aim is to describe the position of the compass after it has been rotated, elevated and rolled. Clearly the order of the transformations is important because we want to describe the elevation and roll relative to the *local* axes on the compass, and not the axes fixed in space.

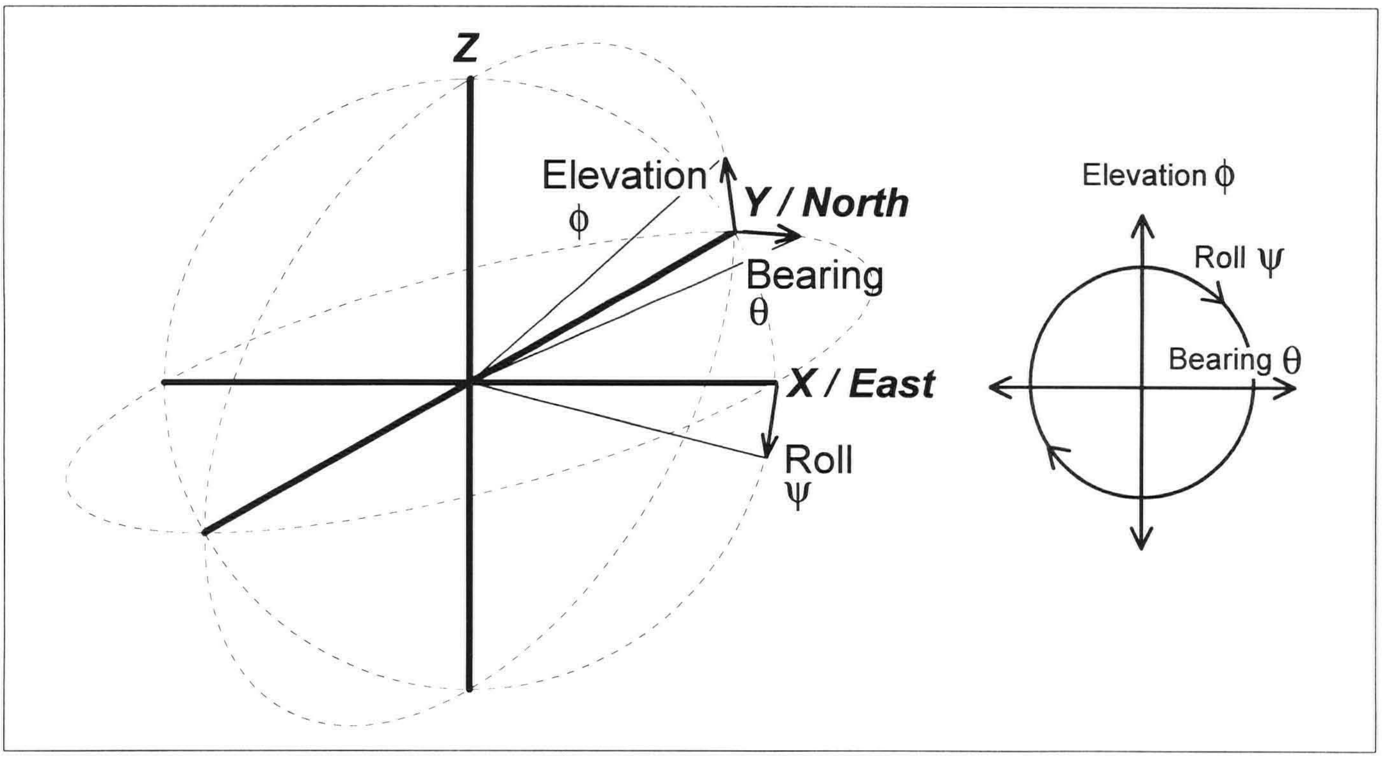


Figure 3. 3-D compass orientation.
The bearing or heading θ is a rotation clockwise about the Z-axis, viewed from above. The elevation or pitch φ is a tilt upwards, about the local X'-axis. The roll ψ is a twist clockwise about the local Y'-axis.

Bearing, Heading, Azimuth, Longitude, Right Ascension
Bearing or heading is the direction measured around the horizon; "north, south, east, west". Azimuth usually refers to true north for astronomical measurements. Longitude and Right Ascension are similar measurements referred to the terrestrial and celestial spheres.

Elevation, Pitch, Altitude, Latitude, Declination
Elevation or pitch is the angle above the horizon, as would be measured by a clinometer. The term 'altitude' is used in civil astronomical observations; Latitude and Declination are similar measurements referred to the terrestrial and celestial spheres.

Roll, Pitch and Yaw
In the flight of an aircraft (or spacecraft, missile, ship) pitch is the angle of the nose-to-tail axis to the horizontal. Roll is a rotation about this axis (one wing up, other down). Yaw is rotation about an axis perpendicular to the plane of the wings.

Tilt – this has no precise meaning.

Angle of declination
The bearing of magnetic north, compared with that of true north. It varies with time.

Angle of dip
This is the angle of the earth's magnetic field lines to the horizontal. It is 90° at the magnetic poles and 67° in London; it varies with time and position.

Figure 4. Glossary of Terms for 3-Dimensional Orientation.

The roll transformation rotates the frame about the Y axis. We want it to act on the Y' axis, so Y and Y' must coincide before we apply **ROLL**. Since the other two transformations both move Y', the roll must be applied first.

The elevation transformation rotates the frame about the X axis, i.e. in the plane $x = 0$. We dont want it to operate in the local $x' = 0$ plane because that might be tilted after the roll operation, but we do want the vertical plane to include the local Y' axis (i.e. we tilt the clinometer in the direction we are pointing it). Thus Y and Y' must coincide before we apply **PITCH** but the other axes need not. Since the roll does not alter Y but the elevation does, we must apply **PITCH** after **ROLL**.

The third operation, a rotation about Z does not cause any problems because it is not referred to any local axes. The transformations must be applied in the order **ROLL, PITCH, HEAD**. The complete expression (with the rightmost transformation applied first) is shown in equation 10 in matrix-vector terms where the overall transformation is written as **TWIST** for brevity.

$$\mathbf{TWIST} = \mathbf{HEAD} \cdot \mathbf{PITCH} \cdot \mathbf{ROLL} = \tag{10}$$

$$\begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{pmatrix} \cdot \begin{pmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{pmatrix} =$$

$$\begin{pmatrix} \cos\theta\cos\psi + \sin\theta\sin\phi\sin\psi & \sin\theta\cos\phi & \cos\theta\sin\psi - \sin\theta\sin\phi\cos\psi \\ -\sin\theta\cos\psi + \cos\theta\sin\phi\sin\psi & \cos\theta\cos\phi & -\sin\theta\sin\psi - \cos\theta\sin\phi\cos\psi \\ -\cos\phi\sin\psi & \sin\phi & \cos\phi\cos\psi \end{pmatrix}$$

Although the expression looks complicated, we can see that if the compass is horizontal (i.e. not pointing up or down, $\phi = 0$) and is level (i.e. not 'rolled', $\psi = 0$) it reduces to **TWIST = HEAD** as we would expect.

MAGNETIC MEASUREMENTS: DERIVING THE BEARING

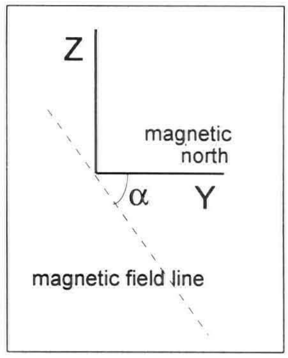


Figure 5. Dip angle of the Earth's magnetic field. This varies with time, but is approximately 67° in London. The flux density is around 48µ T.

Having obtained an expression for the 3-D orientation of the compass/clinometer, the next step is to write down the signals received by the X and Y magnetometers. If the angle of dip of the magnetic field is α (Fig. 5) then these are simply:

$$\begin{aligned} X_M &= \mathbf{TWIST}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \bullet \begin{pmatrix} 0 \\ A_M\cos\alpha \\ -A_M\sin\alpha \end{pmatrix} \\ &= A_M \times \\ &\left\{ \begin{matrix} (-\sin\theta\cos\psi + \cos\theta\sin\phi\sin\psi)\cos\alpha \\ +\cos\phi\sin\psi\sin\alpha \end{matrix} \right\} \end{aligned} \tag{11}$$

and similarly,

$$\begin{aligned} Y_M &= \mathbf{TWIST}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \bullet \begin{pmatrix} 0 \\ A_M\cos\alpha \\ -A_M\sin\alpha \end{pmatrix} \\ &= A_M(\cos\theta\cos\phi\cos\alpha - \sin\phi\sin\alpha) \end{aligned} \tag{12}$$

(X_M, Y_M and Z_M are the signals received on the X, Y and Z magnetic axes. X_G, Y_G and Z_G are the signals received on the gravity axes).

By measuring X_M and Y_M we can derive the apparent compass bearing θ', as shown in equation 2. We can see that in the absence of any elevation or roll we have $\theta' = \theta$. Although the expression is complicated, it is straightforward to demonstrate the effect of a small elevation or roll at certain compass headings. With a bearing of magnetic north (i.e. $\theta = 0$), and the angles expressed in radians, we obtain:

$$\begin{gathered} \tan\theta' = \left(\frac{-X_M}{Y_M} \right) \approx -\psi\tan\alpha \\ \text{but } \theta' << 1 \text{ , so } \theta' \approx -\psi\tan\alpha \end{gathered} \tag{13}$$

which shows that a roll of only 1° (0.035rad) will give rise to a bearing error of 2.4° if the angle of dip α is 67°. In this orientation the elevation does not contribute to the error because the X-axis magnetometer would not detect any field as the elevation was changed. Due east, with $\theta = 90°$ we obtain, in a similar fashion:

$$\Delta\theta' \approx -\phi\tan\alpha \tag{14}$$

where $\Delta\theta'$ is the deviation below 90°. This shows that an elevation of only 1° will give rise to a bearing error of 2.4°. In this orientation the roll does not contribute to the error. Another special case, at $\theta = 45°$, shows that the elevation and roll errors can combine so that a 1° error on each gives rise to an error in θ' as high as 3.3°. Alternatively, the errors can cancel, depending on the sense in which ϕ and ψ are applied.

Having seen that the elevation and roll can contribute significantly to the bearing error in complicated ways, we need to know how to avoid these errors. Clearly one option is to double-gimbal the compass so that it is always horizontal. This is the solution adopted in some range-finder binoculars that have a built-in compass. (We could, of course, combine the gimballing with an optical disc tilt sensor for the clinometer reading). Another option would be to measure the values of elevation and roll with solid-state inclinometers and to use them in an algorithm to derive θ, ϕ and ψ. This has its attractions because we want to know the clinometer reading ϕ in any case.

We might assume, at first thought, that simply measuring the Z-component of the magnetic field would give us the information we require - three axes allowing us to fix the direction of the compass along three axes. This, however, is not so. Three axes describe a cone-shaped locus of possible orientations. (This can be visualised by considering a cardboard box and driving a skewer through opposite vertices. If the skewer is held in a fixed orientation, representing the field lines, then the sides of the box represent the local X, Y and Z vectors. The box can be spun around the skewer without altering the magnitude of the local axes). A fourth measurement is needed in order to fix the direction. This is easily provided by incorporating an inclinometer, which we will describe shortly. It is worth noting that we *could* fix the orientation of the compass with three axes if there were three independent fields we could measure. Unfortunately, gravity and magnetism provide only two fields.

Having established that two further axes are required we have the choice between using a third magnetic axis, with one inclinometer axis; and two inclinometer axes. Using a third (Z) magnetic axis only gives us a limited amount of fresh information, because we could derive the reading from measurements on the X and Y axes, from:

$$Z_M = \sqrt{A_M^{\,2} - X_M^{\,2} - Y_M^{\,2}} \tag{15}$$

but we would be wrong to infer that there was no advantage in using a third magnetic axis.

In the absence of any roll or elevation, (11) and (12) can be solved ratiometrically using (2) without the value of A_M being known. However, in the presence of roll or elevation we cannot obtain the values of $\sin\theta$ and $\cos\theta$ without knowing A_M, which we can derive by measuring three magnetic axes. Another important advantage of using a third magnetic axis is it can warn us if A_M changes significantly during the operation. This would be due to the proximity of ferrous objects that would distort the field. The effect of metal on compass readings is well-known (e.g. Legg, 1995) but here we have a method of sensing the presence of such objects. A third advantage of the Z axis comes when we consider the effects of a gross roll or elevation, which we will do later. And lastly, equation 15 does not, of course, give a unique value for Z_M because we cannot derive the sign.

TILT MEASUREMENTS: DERIVING THE ELEVATION AND ROLL

For the moment, we will assume that the X and Y magnetic axes are supplemented by two inclinometer axes. We begin by assuming that inclinometers are aligned along the local X', Y' and Z' axes and give positive signals when the axes are pointing upwards:

$$X_G = \mathbf{TWIST}\begin{pmatrix}1\\0\\0\end{pmatrix} \bullet \begin{pmatrix}0\\0\\A_G\end{pmatrix} = -A_G\cos\phi\sin\psi \tag{16}$$

and similarly,

$$Y_G = A_G\sin\phi \tag{17}$$

$$Z_G = A_G\cos\phi\cos\psi \tag{18}$$

Manipulating the inclinometer data is straightforward. From the X and Z axes we can derive:

$$\left.\begin{aligned}\tan\psi &= \frac{-X_G}{Z_G}\\ \cos\phi &= \frac{\sqrt{X_G^{\,2}+Z_G^{\,2}}}{A_G}\end{aligned}\right\} \tag{19}$$

or from the X, Y axes we can obtain:

$$\left.\begin{aligned}\sin\psi &= \frac{-X_G}{\sqrt{A_G^{\,2}-Y_G^{\,2}}}\\ \sin\phi &= \frac{Y_G}{A_G}\end{aligned}\right\} \tag{20}$$

If the sensors are calibrated, and do not drift (so we know A_G) then this is sufficient to derive the roll and elevation to within a range of ±90°. If we wanted to use a ratiometric technique that did not depend on absolute measurement of A_G, or if we wanted to allow for a full 360° of elevation and roll, then we would need to use all three axes.

It may seem surprising that we can use two inclinometers to obtain two of the wanted parameters ϕ and ψ, but that with two magnetometers we were not able to derive *any* parameters. This is due to the particular combination of local and fixed axes with which we specify the parameters.

CORRECTING THE MAGNETOMETER DATA

Having derived roll and elevation from two inclinometer sensors, it remains to use this information, together with the angle of dip of the magnetic field, to extract the compass heading θ. Equations (11) and (12) are simultaneous in $\sin\theta$ and $\cos\theta$ but, since the $\sin\theta$ term is absent in (12), they are easily solved to give:

$$\cos\theta = \frac{\dfrac{Y_M}{A_M} + \sin\phi\sin\alpha}{\cos\phi\cos\alpha} \tag{21}$$

$$\sin\theta = \frac{\tan\psi\tan\alpha}{\cos\phi} - \frac{\dfrac{X_M}{A_M} - \dfrac{Y_M}{A_M}\tan\phi\sin\psi}{\cos\psi\cos\alpha} \tag{22}$$

The equations confirm that, as was remarked earlier, we cannot derive a ratiometric solution from Y_M/X_M - we need knowledge of the third axis to derive A_M.

EXTREMES OF ROLL AND ELEVATION

There are a few situations where the manipulation of the above equations needs to be done with care. This is now discussed. The equations so far derived allow us to measure compass bearing in the presence of instrument roll and elevation. Normally we would expect to hold the instrument flat, so the roll would be small, but the elevation - corresponding to the clinometer measurement - may be large. In a surveying application we do not, ultimately, need to know the roll but it is required in order to calculate the bearing.

When the elevation is ±90°, (compass pointing vertically up or down) the term 'bearing' has no meaning. The compass still faces a particular direction, but this changes with the amount of roll and does not correspond to anything useful. If 'bearing' has no meaning, the equations ought to be insoluble, which is indeed the case. We can see from (16) and (18) that with a ±90° elevation X_G and Z_G are both zero, so we cannot derive the roll, which can thus take on any value (what is zero divided by zero?). Without a value for roll, ψ, we cannot use (22) to derive bearing. Further thought shows that, although we could measure the effect of the roll by looking at the X_M and Z_M magnetic axes, this would not give us a unique solution. When the elevation is 90° the orientation can be achieved with many distinct combinations of roll and bearing. In (21), with $\phi = \pm 90°$, the denominator is zero suggesting that $\cos\theta = \infty$, which is not possible. In fact, inspection of (12), from which (21) was derived, shows that the numerator must be zero as well; and that $Y_M = -A_M \sin\alpha$. $\cos\theta$ can then take on any value, from which we infer that the equation is solvable, but that θ is undefined.

A similar situation must be allowed for in (22) should the roll ψ be ±90°. Here we can derive the elevation, but (22) does not allow us to derive $\sin\theta$. In the previous example, this was because the bearing had no meaning. In this case, however, it is because the bearing cannot affect the signal received on the X_M axis when the roll is 90°. Using the Z_M axis allows us to obtain $\sin\theta$ without any problems, using:

$$\sin\theta = \frac{\cot\psi\tan\alpha}{\cos\phi} - \frac{\dfrac{X_M}{A_M} + \dfrac{Y_M}{A_M}\tan\phi\cos\psi}{\sin\psi\cos\alpha} \qquad (23)$$

If the compass is level then Z_M gives no useful information, but for most combinations of elevation and roll either Z_M or X_M can be used as one of the magnetic axes. A suitable algorithm might attempt to derive the bearing from each of these axes.

We can see that the difficulties arise when one of the sensors is aligned normal to the vector it is measuring. Not all axes cause the same problems due, as was noted earlier, to the particular way we choose to specify θ, ϕ and ψ. It remains to check what happens when the magnetometer axes are aligned normal to the magnetic field. If the compass is pointing due north, and level, then the X-axis is normal to the field. If the elevation is 90°-α (about 23° in London) then the Y axis will also be normal to the field. A similar situation occurs for a heading of due east and a roll of α-90°. Clearly there is, for every heading, some combination of roll and elevation which will cause both the X_M and Y_M axes to receive no signal. This observation reinforces the result demonstrated earlier, that roll and elevation could cause an error in the compass bearing of a 2-D device. However, since we already know the values of ϕ and ψ, the lack of signal does not cause a problem; and (21, 22) still have a solution. We can summarise the orientation problems as follows:

Orientation	Problem	Solution
Elevation 90°	No unique values of bearing and roll	None. Bearing has no meaning.
Roll 90°	Cannot derive bearing using X_M, Y_M axes	Don't allow roll of 90°, or use Z_M axis
N/ϕ=23°/ψ=0° E/ϕ=0°/ψ=-23° S/ϕ=-23°/ψ=0° & so on	Neither X_M nor Y_M receives any signal.	Can still solve for θ if we know ϕ and ψ.
Elevation or roll outside ±90°	Incorrect solutions with X_G, Y_G axes.	Use third, Z_G axis

So far, we have considered the problems that occur at specific orientations but it would not be an adequate solution to simply define, in software, the result that 0/0 = 1 because small deviations in the numerator or denominator could lead to vastly different answers. The problem is *not* that bearing cannot be defined at an elevation of 90°. It is that the bearing will be increasingly inaccurate as we approach 90°. It is beyond the scope of this paper to do a detailed error analysis, but this can easily be simulated on a computer and used to predict the performance of a compass module. The specifications for commercial modules (see Gibson, 1996b) do indicate that bearing becomes inaccurate at high degrees of tilt.

SUMMARY

For a 2-D compass we discussed the sources of measurement error, which included quantisation, gain and axis alignment. We demonstrated a simple algorithm which, by using an inverse tangent function, would allow us to derive the compass heading. We showed that a 2-D compass could suffer from errors if it were not held accurately horizontal. This could be achieved by double-gimballing the compass. The gimbal axes could be used to derive elevation and tilt converting the device into a compass/clinometer.

An alternative to gimballing would be to use electronic tilt sensors in a 3-D configuration in order to derive elevation and roll information, and to use this to correct the compass bearing. Unless we knew the magnitude of the magnetic field this would require a third magnetic axis to derive it. We saw that other advantages of a third magnetic axis included the ability to detect errors caused by the proximity of ferrous objects, and the ability to work at high degrees of roll. (This would not normally be required in a caving application).

The use of a third tilt axis, bringing the total to six axes, would allow us to work with values of tilt outside ±90°. Again, this is not necessary for cave surveying, but would be useful in a general application such as borehole orientation.

Equations for the derivation of bearing and elevation using four, five or six axes were presented, together with a general rotation matrix for defining the orientation of the device. The possible reduction in accuracy at high tilt was mentioned, but not analysed.

REFERENCES

Gibson, D, 1996a. Electronics in Surveying. *Journal of the BCRA Cave Radio and Electronics Group,* 23, March 1996.

Gibson, D, 1996b. Review: Electronic Compass Modules. *Journal of the BCRA Cave Radio and Electronics Group,* 23, March 1996.

Legg, A, 1995. Magnetic Effects of Caving Equipment on a Compass. *Compass Points,* 8 (BCRA Cave Surveying Group), June 1995.

Telford, W M, Geldart, L P, Sheriff, R E and Keys, D A, 1976. *Applied Geophysics.* Cambridge University Press (ISBN 0 521 29146 1).

CAVE AND KARST SCIENCE Vol. 23, No.2, October 1996
Transactions of the British Cave Research Association

How accurate is radio-location?

David GIBSON
Department of Electronic and Electrical Engineering, University of Leeds, LEEDS, LS2 9JT

Abstract: The accuracy of a conventional cave survey, constructed using compass, clinometer and tape, and the treatment of the associated surveying errors has been well-discussed. Such surveys are sometimes "corrected" by means of radio-location; but the accuracy of radio-location techniques has not been widely debated. Properly understood, radio-location errors can be subjected to the same treatments as other surveying results. As well as the measurement errors of distance and angle, radio-location accuracy may be affected by the use, in conditions where it is not valid, of the traditional 'quasi-static' model of the field lines.

INTRODUCTION

This paper reviews the sources of error inherent in the use of a radio-location beacon, but it will not attempt to quantify them. In that sense it will not directly answer the question posed in the title of the paper; instead, the intention is to bring the sources of error to the notice of cave surveyors and to encourage a theoretical and practical evaluation of radio-location errors which have, hitherto, not been widely discussed. The subject has received sparse attention in BCRA *Transactions* although Brooks and Ellis (1956) show that attempts in using radio-location for verifying cave surveys go back at least forty years. In the U.S.A. several cavers have done detailed practical studies but this work has not been widely published, nor widely disseminated in the UK.

RADIO-LOCATION TECHNIQUE

Radio-location using an induction loop is, by now, a standard procedure and need not be explained in detail here. A definitive description of the technique was given in *Surveying Caves* by Glover (1976). More recently, Bedford (1993) outlined the technique and presented (with circuit diagram and constructional notes) the electronic beacon previously designed by France and Mackin.

Essentially, a horizontal transmitter loop (vertical magnetic dipole) is placed underground and the point on the surface immediately above this is located using a receiver loop. At this "ground-zero" point the magnetic field lines from the transmitter are vertical so a vertical loop (i.e. with its axis horizontal) will pick up no signal because no field lines "cut" the loop. The ground-zero point is confirmed by holding the loop vertical, spinning it about a vertical axis, and confirming that there is no orientation where a signal can be detected. To locate ground-zero from another location the vertical receiver loop is rotated to give the direction of minimum signal, and a bearing taken along the plane of the loop. A series of at least three widely spaced bearings should, in theory, intersect exactly. In practice the bearings allow the surveyor to construct a "polygon of confusion" which describes a region, on the surface of the earth, in which ground-zero is likely to occur.

The depth of the underground point can be determined in two ways. With suitable equipment the most straightforward method is, perhaps, to measure the flux density (say B_0) at ground zero and to compare this with the signal (B_1) a short distance (y) above this. Using the *ratio* of these readings provides a convenient way of calculating the depth, d, without needing to know the transmitter power or absolute gain of the receiver. The inverse cube law, which describes the change of flux density with distance leads directly to:

$$\frac{y}{d} = \left(\sqrt[3]{\frac{B_0}{B_1}}\right) - 1 \qquad (1)$$

This method, which could be termed 'depth by signal-strength' (DSS) is used in some commercial radio-location equipment, but most amateur designs have used a different method based on measurements of the field angle. The reasons for this are not entirely clear; it may be due to the nature of some amateur amplifier designs which make direct readings of field strength difficult to obtain; or it may be that a discrete design, based on the electronic components which were available 20 years ago, would have been complicated. The measurement of depth by signal-strength has been discussed in caving literature, but its dismissal may have been due to a lack of insight into ratiometric techniques. Using a ratiometric technique instead of trying to relate absolute signal strength to depth avoids the perceived problems in maintaining the transmitter power and the receiver gain. Currently available commercial equipment uses a micro-controller to allow direct "real-time" readings of depth to be obtained. Brian Pease (1995) is currently experimenting with a DSS device for cavers, but this does not use a ratiometric technique.

The more common method of depth determination is to measure the angle of the field lines. Away from ground-zero the magnetic field lines are not vertical. By measuring the angle of the field to the ground (α), and knowing the distance to the ground-zero point (x), the depth of the transmitter (d) can be calculated (Fig. 1). This method assumes that the field lines obey the parametric equations for a traditional "bar magnet". The formula is:

$$\frac{x}{d} = \frac{\sqrt{\left(8 + 9\tan^2\alpha\right)} - 3\tan\alpha}{2} \qquad (2)$$

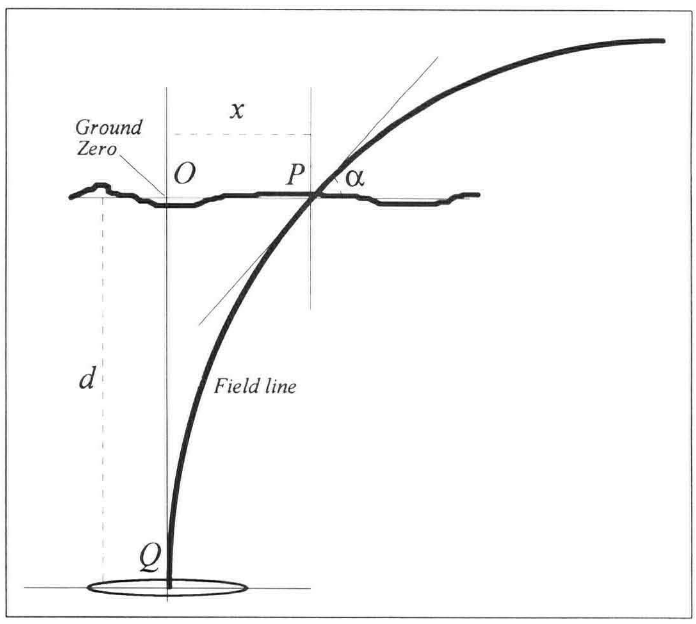


Figure 1. Determining depth by measuring the angle of the field lines.

This result is straightforward to derive, and has been familiar to cavers for many years now. It is quoted by Mixon and Blenz (1964), also by Glover (1976) in *Surveying Caves*. For a simple derivation see, for example, Lee in an appendix to Glover (1973) or, more recently, see Gibson (1994).

A convenient technique for depth estimation is to find the distance x at which the field lines lie at 45° to the ground. The formula then indicates that $x/d \approx 0.56$, so the depth is approximately twice the distance x. Another technique would be to find the distance x at which the field lines were at 18.4°, for which $x/d = 1$.

MEASUREMENT ERRORS

Clearly there is scope for errors of measurement to have a significant effect. Most of the sources of error affect the depth measurement more than they affect the location of ground-zero. The accuracy of a position fix also depends, of course, on the accuracy of the surface survey. Ideally several bearings would be taken, in order to locate ground-zero as accurately as possible. Then field-angle measurements at varying distances would be plotted, and used to obtain a best-fit curve from which the depth would be determined. In practice, cavers might only make one or two measurements but, if this is the case, the confidence of the result must be called into question. Glover (1976) demonstrated various graphical methods of converting α and x into depth. His graphs show how small errors in reading can lead to large errors in depth. If $\alpha = 80°$, for example, then a 1° increase in α corresponds to an decrease in x/d from 0.117 to 0.105, which is 10%. At $\alpha = 45°$ the change is only 2.5%. Mixon and Blenz (1964) also discussed angular errors in their paper.

Measuring the angle of the field lines on the surface requires the surveyor to accurately sight on the ground-zero point. As he adjusts for the null position, by tilting the receiver loop, he must ensure that it remains pointing towards ground-zero. Obtaining an accurate null, and accurately measuring the angle of the loop are crucial aspects of the technique; and obtaining a good null is not always easy. There is a *secondary field* effect, to be described later, which builds up rapidly away from ground-zero and makes it increasingly difficult to get a deep null as the angle of the field lines, α, decreases. Depth measurements should, ideally, be made with α from 40° to 50°.

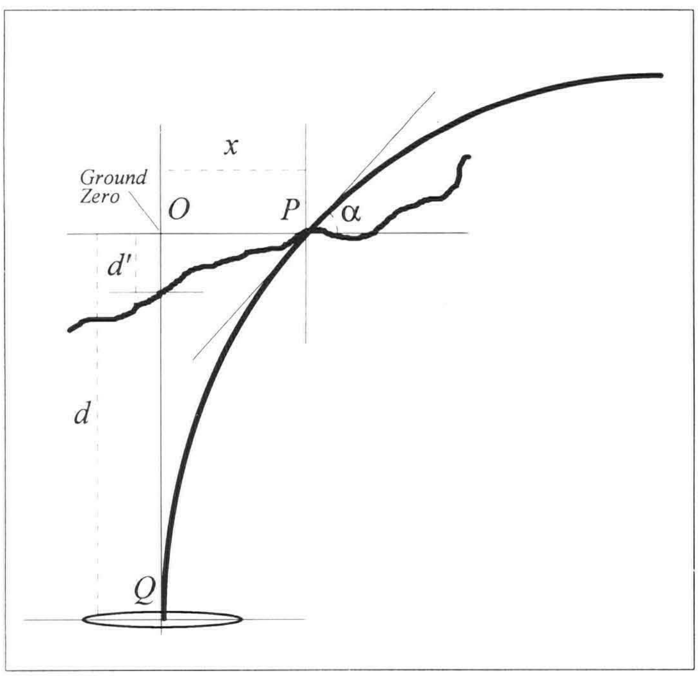


Figure 2. Rough Terrain.
Distance d' must be determined in addition to x and α.

If a ratiometric DSS technique is used there is a limitation caused by how far above the ground the signal strength can be measured. Here, too, a secondary field effect can contribute to the error by causing the above-ground field to behave in a different way to the underground field.

The underground transmitter must be set up as accurately horizontal as possible. If the transmitter is only levelled to 5° the axial field line will be displaced by 8.7% of the depth (i.e. tan 5°). This field line will not be vertical but, from figure 1 and equation 2, it can be shown that for a small tilt (say < 10°) the field line which is vertical as it leaves the ground will be displaced by a third of the distance to the axial field line. Thus, with a 5° tilt the apparent ground-zero point will move by about 3% of the depth. Significantly, if the transmitter loop is not completely horizontal there will not be a field line which *remains* vertical as it leaves the ground. This could cause the null to be less sharp since there will always be some lines cutting the loop. In practice the loop can be levelled to better than 5°, but a spirit level is essential, as is a neatly wound induction loop. The terrain can be a source of measurement error because the ground-zero point may not coincide with the surface of the ground (Fig. 2). If the ground is sloping then equation 2 can still be applied, but x must be the true *horizontal* distance to ground zero, and d must be measured from the altitude of the field point. Distance d' must be determined by surveying. Another source of error in addition to the obvious "sighting" errors is that it is possible to detect a false ground-zero in particularly rough terrain, especially if the estimation of the surface location is tenuous to begin with (Reid, 1990).

FIELD LINE DISTORTION

There is another source of error, potentially far more serious than the measurement errors described above. It is caused by the magnetic field lines departing from the supposed "bar magnet" shape. There are several reasons for this.

i) The receiver loop may be too close to the transmitter

Unless the receiver is far enough away for the transmitter to look like a point source, the field lines will not be of the simple 'bar magnet' shape which is usually derived by considering a quasi-static field from a point-source dipole. In practice this means around five diameters, and this will not normally be a problem unless a large transmitter or receiver is used. For example, a 2m loop requires a depth of at least 10m in order to get an accurate reading. With a smaller loop, good results can probably be obtained closer than five diameters because a larger margin of error can be tolerated. Not only are the field lines distorted in the immediate vicinity of the transmitter loop, but the familiar inverse cube law breaks down too, so equation 1 cannot be used for depth estimation. It is possible to derive an expression for the field from a loop of finite extent but it is complicated and therefore of limited application. One procedure is to integrate the standard expression for 'retarded potential' over a suitably defined current density distribution. Mixon and Blenz quote a result; and show that it reduces to the simpler "bar magnet" field when the field point is at a large distance from the loop.

ii) The field lines will be distorted by magnetic rocks

The distortion of field lines is exploited by geophysicists and archaeologists, who use magnetometers as surveying tools. Unfortunately, unless a control grid of readings is correlated with an accurate compass and clinometer survey, the extent of the problem will not be known.

iii) Distortion by conductive rock – the "phase" problem

The absence of magnetic rocks and minerals does not imply that the field lines are undistorted because conductive, but non-magnetic, rock can *also* distort the field. Radio-location has to be used with care in areas where there is much mineralisation. This effect is well-known to geophysicists and archaeologists who utilise a magnetic gradiometer to induce a field in conductive rock; the field gradient is then a measure of the distortion of the field lines, and allows the structure of the ground to be determined.

The effect of a magnetic field passing through conductive rock is to introduce eddy currents. This generates a so-called *secondary* magnetic field. This field is out of phase with the primary field and therefore leads to elliptical polarisation which prevents a deep "null" condition from being obtained. The problem was discussed by Drummond (1987a) and Gibson (1993a). It is worse at larger distances. The secondary field is of use to geophysicists, who can use it to measure conductivity by a non-contact means. (Pease, 1991, 1995).

iv) The "Transition Zone" problem

The field from an induction loop can be divided into two regions. The *near-field* (or induction field) predominates at distances less than $\lambda/2\pi$ (λ is wavelength). The *far-field* (or radiation field) predominates at distances greater than this. The two fields have very different properties. For a large distance either side of $\lambda/2\pi$ there is a *transition zone* where the field gradually changes from the induction "bar magnet" shape to concentric circles which do not intersect the origin. The inference is obvious – within the transition zone the field lines will not be the simple "bar-magnet" shape which is predicted by the "quasi-static" model.

THE "POLARISATION" PROBLEM

The "transition zone" and "phase" problems can be discussed together as a "polarisation" problem. One or other of the effects have been observed by a number of cavers, though the effects are not always attributed to the correct causes. For example, a comment like "*we could not find a null because the signal was so strong*" (Williams and Todd in *Caves and Caving*, 35, Spring 1987) should probably be attributed to the predominance of the secondary field. The transition zone is centred on $\lambda/2\pi$ and this might be expected to be large at the low frequencies used for radio-location. However, the crucially important point is that the wavelength *in the rock* is much less than this. The transition zone moves inward to δ (and the wavelength to $2\pi\delta$) where δ is the skin depth, given by:

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}} \qquad (3)$$

Here ω is 2π x frequency [Hz], μ is magnetic permeability of the rock [H/m], and σ is electrical conductivity [Ω^{-1}/m]. Strictly speaking this expression is only true for a "good" conductor, but it applies to most rock. (Rock is a "good" conductor by the mathematical definition of $\sigma/\omega\varepsilon >> 1$ unless it is very dry and the frequency is high). Note that the skin depth does *not* describe a physical skin in which the signals are constrained to lie. The signals can and do penetrate further than the skin depth, which is simply a useful mathematical "figure of merit" for the rock. A derivation of the above result, with specific reference to "good" and "bad" conductors, was given by Gibson (1996).

Skin depth can vary from a metre or two to several hundred metres for the range of frequencies and rock types encountered by cavers. It is quite conceivable that a radio-location beacon could be operating at depths comparable with the skin depth and where the transition zone effects would be significant. At this distance, secondary fields would also be significant. Interestingly, the optimum depth for communications (as opposed to radio-location) may be around three skin depths (Gibson, 1993b, 1994).

The subject of radio wave propagation through rock has been well-studied, although the results have often been presented in a mathematical form which is not easy for non-mathematicians to interpret. Steven Shope (1991) has summarised some previous results and presented them graphically, showing how the direction of the field lines at the surface depends on the skin depth. One of these graphs was reproduced by Bedford (1993). Shope's graphs are extremely significant because they show that in some circumstances the result given by (2) can be very much in error. It is intended that this will be the subject of further study by the author. It is worth pointing out that it is not only the field angle α which departs from simple "bar magnet" theory; the $1/D^3$ rule for flux density also breaks down in the transition zone so, under these conditions, equation 1 cannot be used for depth determination either. There is some indication (Pease, pers. comm.) that, under these conditions, DSS gives rise to an over-estimation of depth, whereas field-angle measurement gives rise to an under-estimation.

ACCURACY OF RADIO-LOCATION

The measurement problems can be quantified and used to make an estimate of the accuracy, which could easily be 5-10% for depth, and several metres for ground zero. The polarisation problems are less easy to quantify. Depth determination starts to fail if a good null cannot be obtained, eventually failing completely. In these circumstances a ground-zero location can often still be performed. This only starts to fail if the rock is anisotropic, or if the ground is inhomogenous in a radially non-symmetric way (e.g. the antenna is located close to a fault-line or to one side of a large cavity).

AVOIDING THE PROBLEMS

The problem of the transition zone is lessened considerably by using a very low frequency, because the zone is further away, and because the secondary fields have a lower magnitude. The France/Mackin beacon operates at 874Hz; several US designs operate at 3496Hz. Radio-location at these frequencies is likely to be more successful than if it is done using carrier-based speech systems; common frequencies for which are around 27, 87, 115 and 185kHz; see Bedford (1994).

PRACTICAL MEASUREMENTS

Ian Drummond (1987b) has described some experiments which he, and others, performed in Lechuguilla Cave in New Mexico. Amongst them was a series of radio-locations along a passage at a depth of up to 210m. The purpose was to see if the magnetic field was well behaved, and if it diverged symmetrically from the null point. Plotting the data (and re-surveying part of the cave to check for errors) showed that the field was badly distorted in one area. This was attributed to mineralisation of a particular cross-rift. The experiments confirmed the wisdom of performing a series of locations to provide a control grid for a survey, rather than relying on one single point at the far end of the cave to check the survey. Drummond also found that the sharpness of the nulls depended on the orientation of the antenna. The precision of the location on the surface was much better along the passage than at right angles to it. This may well be a secondary field effect, but Drummond has noticed a similar effect on other occasions and suggests (pers. comm. quoted in Gibson 1993b) that it could be an anisotropic characteristic of the rock.

In the UK, members of the BCRA's Cave Radio Group are currently re-surveying Kingsdale Master Cave and performing a series of radio-location fixes. In addition to providing some simple tests of the accuracy of the radio-location, this will pave the way for a set of experiments, at different frequencies, which will attempt to verify Shope's graphs.

An observation arising from experiments in several countries, is that the UK suffers comparatively badly from high rock conductivity and high levels of background interference. The inferences are that polarisation effects are likely to be worse, and that nulls are likely to be less sharp. The Cave Radio Group has demonstrated that radios which penetrate well in the US do not operate so well in the UK.

There will always be errors associated with the measurements made using radio-location beacons, and a proper understanding of them is essential. Cavers who have used radio-location beacons have sometimes misunderstood the operation of the device – Williams and Todd's comment was quoted earlier. Other cavers have (pers. comm.) taken bearings of ground zero and, because the readings have intersected to give a triangle of error, the cavers have deduced that the beacon "was not working properly". Another common mistake is to assume, without justification, that the results obtained by radio-location are 100% accurate. Statements such as "we fixed the position by radio-location" suggest a misplaced confidence that the technique has an unfailing accuracy.

SUMMARY

Radio-location works best at very low frequencies (below a few kHz) and over distances which are short compared to the skin depth, but large compared to the size of the loop. Its accuracy is affected not only by measurement error, but by factors which are difficult to predict, such as distortion of the field lines. To use radio-location to best advantage users must understand the nature of the errors; they must know how to minimise them, and should know how to deal with uneven terrain.

This paper was intended to make users aware of the possible inaccuracies of radio-location, rather than to ascribe precise figures to the sources of error. Occasional tests of accuracy have been made but not widely reported; and it is hoped that this paper will encourage further discussion of both theoretical and practical aspects of the technique.

ACKNOWLEDGEMENTS

I would like to thank those members of the BCRA's Cave Surveying Group (Olly Betts, Bryan Ellis, Stuart France and Wookey) who discussed an early draft of this paper with me; also Brian Pease, who discussed the results of sophisticated signal measurements with a radio-location beacon; and especially Ian Drummond, who made many useful comments and reminded me of the work done by Bob Buecher, Frank Reid and others.

REFERENCES

The list of references contains material specifically oriented towards cavers, but the techniques are well-covered in geophysics and electromagnetics textbooks, as well as in various caving club publications.

Bedford, M, 1993. An Introduction to Radio-Location. *Journal of the BCRA Cave Radio and Electronics Group,* 14, Dec. 1993, 16-18.

Bedford, M, 1994. A Directory of Cave Radio Designs. *Journal of the BCRA Cave Radio and Electronics Group,* 18, Dec. 1994, 3-4.

Brooks, N and Ellis, B, 1956. An Independent Check on the Survey of OFD. *South Wales Caving Club Newsletter,* 16.

Drummond, I, 1987a. Magnetic Moments #5: The Phase Problem. *Speleonics* 7, Spring 1987, 11-12.

Drummond, I, 1987b. A Cave Radio in the Field – Summer 87. *Speleonics,* 9, Winter 1987, 7-8.

Gibson, D, 1993a. An Introduction to Cave Radio. *Journal of the BCRA Cave Radio and Electronics Group,* 12, June 1993, 24-26.

Gibson, D, 1993b. Penetration of Magnetic Fields Underground – Part 3. *Journal of the BCRA Cave Radio and Electronics Group,* 14, Dec 1993, 21-25.

Gibson, D, 1994. Penetration of Magnetic Fields Underground – Appendices. *Journal of the BCRA Cave Radio and Electronics Group,* 15, March 1994, 19-21.

Gibson, D, 1996. Cave Radio notebook – 10: Skin Depth. *Journal of the BCRA Cave Radio and Electronics Group,* 24, June 1996, 17-18.

Glover, R R, 1973. Cave Depth Measurement by Magnetic Induction. *Lancaster Uni. Speleo. Soc. Journal,* 1(3), 66-72.

Glover, R R, 1976. Cave Surveying by Magnetic Induction *in* Ellis, B (ed.), *Surveying Caves*, BCRA.

Mixon, W and Blenz, R, 1964. Locating an Underground Transmitter by Surface Measurements. *The Windy City Speleonews,* 4(6) 47-53. [also: reprint in Speleo Digest 1964].

Pease, B, 1991. Measuring Ground Conductivity with a Cave Radio. *Speleonics,* 16, May 1991, 4-6.

Pease, B, 1995. Letters, *Journal of the BCRA Cave Radio and Electronics Group,* 21, September, 29.

Reid, F, 1990. False Center Found in Unusual Location. *Speleonics,* 15, Oct. 1990, 15.

Shope, S, 1991. A Theoretical Model of Radio Location. *NSS Bulletin,* Vol. 53,

CAVE AND KARST SCIENCE Vol. 23, No.2, October 1996
Transactions of the British Cave Research Association

Addendum

A short section was missing from the text compilation provided for the Yangtze Gorges Expedition 1994 Report (Cave and Karst Science, Volume 22, Number 2). The omission was noticed neither by the Report compiler nor by the editors. The missing section is reproduced below, together with a reduced version of the original Figure 7, to which the text refers.

Downstream Xio Zhai Tien Ken

This is a classic through trip, leaving the daylight in Xio Zhai Tien Ken and exiting into the Mie Gong He Gorge, 4.5km away. It is currently the deepest cave in China, measured from the lowest edge of the doline where the path descends.

In dry conditions, most of the water from upstream Xio Zhai Tien Ken is diverted into a tunnel and thence to a hydro-electric power plant. The river in downstream Xio Zhai Tien Ken dries up, leaving static pools among slippery rounded boulders. In flood conditions a torrent of water sweeps over the dam and across the floor of the doline into the downstream cave. There would be little chance of escape for explorers in these circumstances.

Large slippery boulders lead to the head of a 14m shaft, which drops immediately into a deep blue lake. This is the first of many such lakes, whose combined length is nearly 2km. The water is cold and static. 30m across the lake and with only a narrow lip out of the water, is a 65m shaft. 7m down the shaft is a ledge, and a little beyond that the drop is free-hanging into a bigger lake. About 5m away there is a ledge out of the water, but the way on involves wading about 90m, to land between large rounded boulders. Looking back there is a superb view up the entrance shafts, out into the doline and to the skyline, more than 800m above. In the cave, the passage is approximately 20m wide and at least 90m high, with very smooth walls. This height is maintained throughout the cave and, generally, it is not possible to see the roof with normal caving lights.

Climbing over and around boulders with deep water-filled holes between, leads to another lake. The passage here is narrower, but just too wide to bridge across. A succession of lakes, mostly about 60m long, alternate with wider, cobble- floored passage. Huge logs, worn smooth by the force of flood water, are jammed among the boulders. There may be a high level passage up on the left where the cave turns southeast for a short section (Fig. 7). This may be a former upstream continuation of downstream Xio Zhai Tien Ken, active before the upstream cave was captured.

Beyond this point several small inlets enter the cave, and a longer (140m) canal drains down between boulders. By keeping up high, a ledge is reached, from which a 10m abseil drops back to the water. The cave continues with a succession of long lakes, some separated by only a boulder. 1.5km from the entrance is a larger inlet, and the water in the cave is no longer static.

2.5km from the entrance the most noticeable feature is a very strong draught, which sends ripples across the surface of the lakes, especially in The Wind Tunnel, a 300m-long lake. The draught is lost where the passage starts to cut down, and a high level passage to Mie Gong He Dong could exist here.

A 3m climb, for which a rope is useful, drops into a pool that is held back by boulders before dropping noisily over a waterfall. 10.5m and 4.5m shafts can be rigged alongside the waterfall, largely out of the force of the water. A handline is advised on the 2m-high cascade immediately below. The passage is now at its narrowest, less than 1m wide, and this section

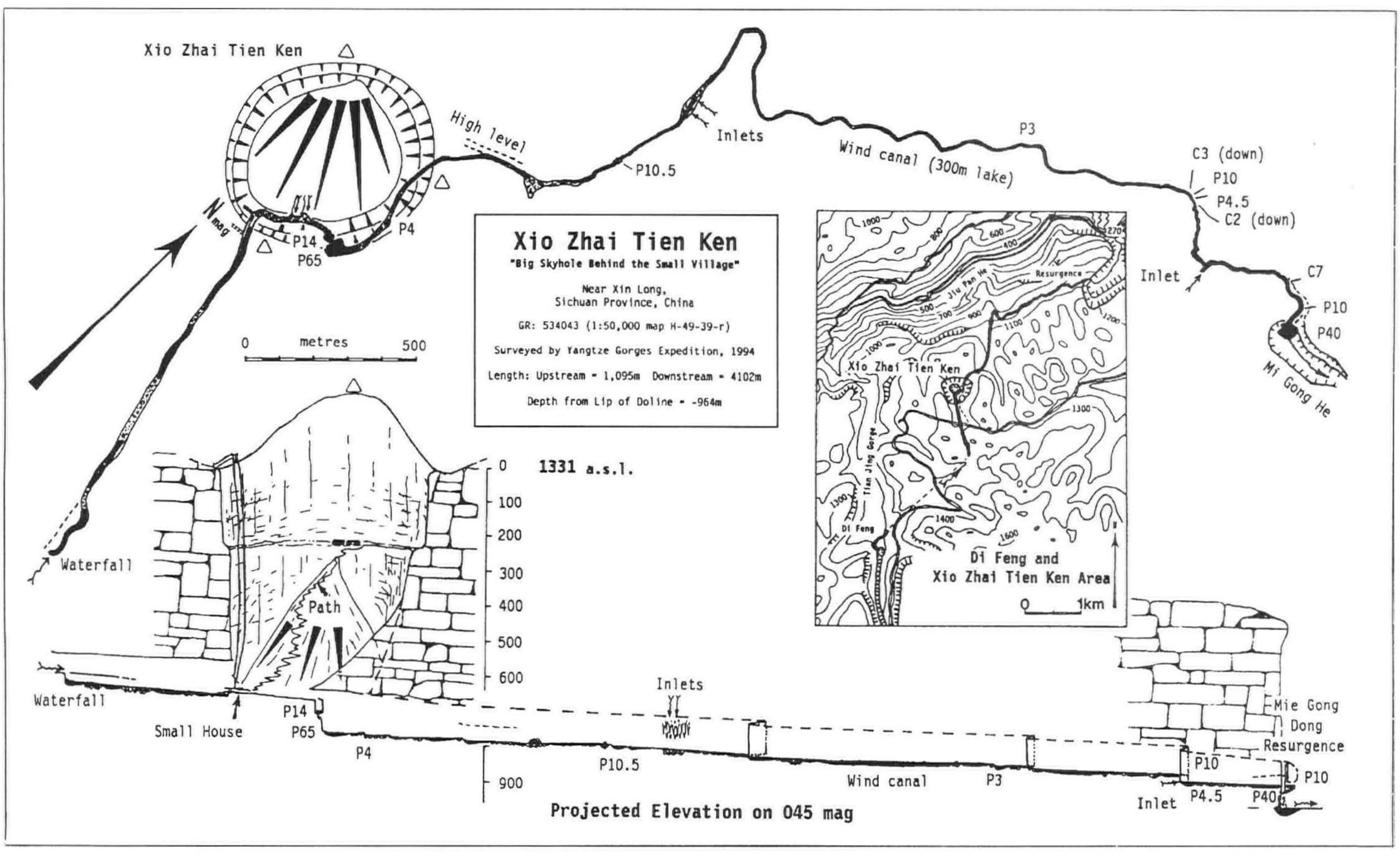


Figure 7. Xio Zhai Tien Ken survey.

could become impassable in wet conditions. Large tree branches are jammed across the passage, almost blocking it. From this point daylight can be seen faintly, high in the roof, but it is some distance before light reaches the floor, and the exit from the cave is still nearly three quarters of a kilometre away.

An inlet enters, the passage widens again to 2 or 3m, and more lakes lead to a 7m climb down through boulders, for which a rope is useful. A large boulder holds back a lake, on the other side of which is a 10m shaft. The lake at the bottom meanders around a corner in the high canyon passage and leads to the lip of a 40m shaft into the Mie Gong He Gorge. It is a spectacular site, with the walls of the gorge overhanging by about 30m some 700m above and another waterfall drifting lazily down the opposite wall.

After descending the last shaft and swimming the lake at the bottom, two more lakes in the Mie Gong He Gorge must be crossed. An indistinct path on the right leads up, through dense vegetation, to the village of Yitz, and from here it is another 3km to the road from Xin Long.

VOLUME 21 (3), PAGE 8

Onur Ozbek has asked us to note that he no longer has any relation with the "Cave Research Association" and that his correct address is:

Natural History Research Association,
11.Sk. No: 29/11,
Bahceliever,
06490,
Ankara,
Turkey

CAVE AND KARST SCIENCE Vol. 22, No. 3, December 1995
Transactions of the British Cave Research Association

Forum

Readers are invited to offer thesis abstracts, review articles, scientific notes, comments on previously published papers and discussions of general interest for publication in the Forum of Cave and Karst Science.

THESIS ABSTRACTS

KING, H.K., 1996

The U-series Dating and Geochemistry of Flowstones from the Great Orme Mine

Unpublished PhD thesis, University of Liverpool, P.O.Box 147, Liverpool, L69 3BX. UK
[Available from the University]

Seven calcium carbonate ($CaCO_3$) flowstones were collected from various sites within the Great Orme Bronze Age copper mine in order to establish an ante-quem age for their underlying ancient mining back-fill. The age of each flowstone was obtained by analysing a succession of samples using the U-series disequilibrium method and extrapolating a best fit line through the dates to the base. Five of the flowstones were of calcite morphology (determined by X-ray diffraction) and each contained c. 1ppm U. The dates (2.38 ± 0.22, 1.70 ± 0.63, 3.93 ± 0.40, 3.52 ± 0.94 and 2.83 ± 1.05 ka BP) provide additional weight to the evidence for the Bronze Age origins of the mine and the last three are comparable to ^{14}C dates (3.63 ± 0.21, 3.41 ±0.05 and 3.115 ± 0.245 ka BP, respectively) from adjacent wood and bone samples. The other two flowstones were aragonite and contained c.70ppm U. The dates (1.53 ± 0.00 and 0.1 ± 0.0 ka BP) had negligible statistical errors due to the high U concentrations. The latter date was obtained from a flowstone deposited in an area abandoned in the 1850s and shows that, with enough U present, very young ages may be determined from secondary carbonates.

Microscope analysis of thin sections showed considerable variation in crystal habit and structure both between flowstones and within individual specimens. Interestingly, one of the aragonite flowstones had a thin layer of calcite forming its base, a distinct hiatus was seen between the two morphologies and the calcite was clearly not the result of aragonite inversion. Trace element analysis revealed a linear relationship between the concentrations of Mg and Sr and the carbonate morphology, with the calcite exhibiting a higher Mg and a lower Sr concentration than the aragonite.

Stable isotope analysis of four flowstones with varying physical features revealed that the $CaCO_3$ was not deposited in isotopic equilibrium with its parent waters. This suggests that the deposition was influenced by a process such as evaporation or rapid outgassing of CO_2. However, the flowstones exhibited similar mean $\delta^{18}O$ values (-4.59 ± 0.23, -4.51 ± 0.27, -4.34 ± 0.22 and -4.03 ± 0.24 ‰VPDB) which relate to the present day climatic conditions. It is thought that the relatively consistent $\delta^{18}O$ of the flowstones represent a relatively consistent deposition mechanism. Variations in $\delta^{13}C$ between the flowstones are likely to be due to the variations in the thickness and lithology of the overlying rock through which the parent waters percolated. A difference in the fractionation of ^{18}O and ^{13}C is observed between the calcite and aragonite which corresponds to theoretical and experimental estimations of the effect.

KNEZ, M., 1996

The influence of bedding planes on the development of karst caves (in Slovenian and with an English summary and abstract)

PhD thesis, University of Ljubljana.
[Available from Karst Research Institute, Centre for Scientific Research of the Slovene Academy of Sciences and Arts, Titov trg 2, SI-6230 Postojna, Slovenia]

There have been much less researchers looking for the initial water ways in karst along the bedding-planes than those who deduced the origin of cave channels from tectonic structures. The aim of my research was to focus scientific attention on the sphere where the answers within the sedimentology might be expected. The study identified that the basic idea of bedding-plane importance at the initiation of cave channels was correct but also, that the interrelation is different from how it had been supposed. Single lithological, petrological or stratigraphical parameters of the inception are only partly known, or merely guessed. My research threw light on the problem of initial channels met in Velika dolina in Škocjanske jame. Cave passages, or their fragments and other traces of the underground karstification do not appear scattered at random on the walls but they are obviously gathered along a small number of so-called bedding-planes.

The basic working method was to locate the phreatic channels or their fragments, to sample and microscope those parts of the layers adjacent to a bedding-plane. Somewhere a whole layer was considered. Other methods were: regional distribution of caves, photographing, inventarisation and classification of speleogens and complexometry, the latter providing the purity of limestones.

The original channels are practically gathered along only three formative bedding-planes (out of 62 measured); their close vicinity differs from the others in several important properties: typically damaged rock, higher level of calcium carbonate, smaller porosity and others. Consequently the mentioned concordance cannot possibly be only apparent.

From the lithological point of view, I got neither substantial argument nor explanation for selective karstification. However, it was identified that at least in respect of a concrete example from Velika dolina, the inception started along interbedded slides that without doubt pushed the beds aside leaving an interval.

LEAN, C.M.B., 1995

Geomagnetic Palaeosecular Variation Recorded In North And Central American Speleothems

Unpublished Ph.D. thesis, University of Liverpool, P.O.Box 147, Liverpool, L69 3BX. UK
[Available from the University]

The aim of this project was to collect samples of stalagmites from Northern and Central America in order to produce records of the palaeosecular variation of the earth's magnetic field. Two stalagmites were sampled from Western Canada and ten from Mexico and Guatemala which could be compared with contemporaneous stalagmite records from these areas (Latham, 1981; Latham et al, 1982; 1986; 1987; 1989).

The stalagmites were generally weakly magnetised but remanence directions were stable upon stepwise thermal and alternating-field demagnetisation. Consistency in directions recorded between central and corresponding lateral sub-samples within two stalagmites (MSC2 from Canada and CP1 from Guatemala) inferred that any depositional errors caused by surface effects were less than the measurement errors. Grain size analysis showed the presence of a fine-grained magnetic fraction (0.01 - 0.1∝m) sourced from the cave drip-waters (either by direct deposition or by chemical precipitation) and a coarser magnetic fraction (0.01 - >10∝m) sourced from the flood-borne detritus. The latter source was dominant in stalagmites which were regularly inundated with water. The type of magnetic mineral present was determined by the geology of the catchment area; magnetite dominated in the Vancouver Island stalagmites, titanomagnetite in the Mexican stalagmites and haematite in the Guatemalan stalagmite.

Uranium-series dating of samples was hindered by the young ages of many of the samples, by low uranium concentrations and by the presence of allogenic thorium. If significant amounts of allogenic thorium were present, a sample age could be calculated based on an estimate of the initial thorium ratio ($[^{230}Th/^{232}Th]_0$). Analysis of samples from Sumidero Recuerdo in Mexico, however, suggested that this ratio is not constant with time and may vary by a factor of two over approximately 1700 years. Due to these imprecisions many dates were out of stratigraphic sequence and age estimates were made assuming constant growth rates, except where growth had ceased for a finite length of time.

Records of sequential change of palaeomagnetic direction were obtained from the Mexican stalagmite SSJ3 and the Canadian stalagmite MSC2. The reliability of the latter record was confirmed by comparison with another Canadian stalagmite record (Latham et al, 1987) and contemporaneous lacustrine records. Other records were disappointing due to poor temporal resolution; each sub-sample represented a period of approximately 1000 years in Mexican stalagmites SSJ2 and SSJ4. Such slow growth rates are insufficient for the resolution of secular variation features with periods of less than 2000 years and are only suitable to gain information about the nature of long-term secular variations, for example the far-sided virtual geomagnetic poles and low inclinations predominant throughout the Holocene in Southern Mexico.

The existence of matching contemporaneous stalagmite records of secular variation together with the demonstrated lack of depositional inclination errors is encouraging, despite the sometimes "hit or miss" aspects of sample selection. Nevertheless it has been proved that speleothem records have the potential to complement the existing archaeomagnetic, lava and lacustrine data.

OPENSHAW, S.J., 1996

Palaeosecular Variation Observed In Speleothems From Western China And Northern Spain

Unpublished PhD thesis, University of Liverpool, P.O.Box 147, Liverpool, L69 3BX. UK
[Available from the University]

This study has produced records of the palaeosecular variation (PSV) of the earth's magnetic field from speleothems from China and Spain. The ultimate aim of this project was to produce contemporaneous PSV records which would show that speleothems accurately record ambient geomagnetic field behaviour. From Sichuan Province, China, five speleothems were collected of which four were studied for their records of PSV. Eight Spanish speleothems from the Cantabrian coast were collected but their weak magnetisation allowed only one record of PSV to be produced.

All speleothem sub-samples were weakly magnetised and had, on average, initial intensities of <100 x $10^{-8}Am^2kg^{-1}$. Despite this, the majority of sub-samples were stable during stepwise alternating-field and thermal demagnetisation and each displayed a single component of magnetisation after removal of any secondary overprints. Rock magnetic experiments were hampered by low mineral concentrations but suggested that the remanences of each speleothem were carried by a mixture of multi and single-domain (titano-) magnetite and also by haematite present in significant quantities. The primary method of remanence acquisition appeared to be a depositional remanence sourced from flooding. This was corroborated by a linear relationship between sub-sample intensities and weight % acid insoluble detritus.

A selection of sub-samples from each speleothem were dated using uranium-thorium disequilibrium and alpha spectrometry. For the majority of sub-samples the low concentrations of uranium, high levels of detrital contamination and initially low chemical yields raised the associated dating inaccuracies above the quoted level for alpha spectrometry of 5-10%. Two Spanish speleothems had high uranium concentrations and little, or no, detrital contamination. Percent age errors of these speleothems ranged from 1 to 6%. Comprehensive experiments on the efficiencies of three electrodeposition methods were also undertaken. The most efficient method was found to be a modified version of the Hallstadius method (Hallstadius, 1984) which consistently achieved chemical yields between 40 and 90% for uranium and thorium.

In order to correct more analytically for the presence of detrital contamination, the leachate/leachate method of Schwarcz and Latham (1989) was tested The maximum likelihood estimation data treatment technique (Ludwig and Titterington, 1994) was used to calculate dates from these analyses. Tests on Mexican speleothem SSJ2 gave excellent results allowing a revised dating scheme to be adopted. Tests on some sub-samples from Chinese speleothems were generally unsuccessful due to analytical errors.

The isotope ^{210}Pb was used to date the top surface of one speleothem. A constant growth rate was inferred which was significantly less than that calculated from the ^{230}Th - ^{234}U dating method. This was thought to be due to the former techniques inability to resolve growth rates of periods of less than 200 years.

Despite the dating errors associated with each speleothem the records of PSV compare well with each other and with contemporaneous records from China, Japan and also the UK (for the Spanish record). In addition, agreement with PSV data modelled from observatory records suggested that westward drift of the non-dipole geomagnetic field was predominant during the past 10ka.

RESEARCH FUNDS AND GRANTS

THE BCRA RESEARCH FUND

The British Cave Research Association has established the BCRA Research Fund to promote research into all aspects of speleology in Britain and abroad. Initially, a total of £500 per year will be made available. The aims of the scheme are primarily:

a) To assist in the purchase of consumable items such as water-tracing dyes, sample holders or chemical reagents without which it would be impossible to carry out or complete a research project.
b) To provide funds for travel in association with fieldwork or to visit laboratories which could provide essential facilities.
c) To provide financial support for the preparation of scientific reports. This could cover, for example, the costs of photographic processing, cartographic materials or computing time.
d) To stimulate new research which the BCRA Research Committee considers could contribute significantly to emerging areas of speleology.

The award scheme will not support the salaries of the research worker(s) or assistants, attendance at conferences in Britain or abroad, nor the purchase of personal caving clothing, equipment or vehicles. The applicant must be the principal investigator, and must be a member of the BCRA in order to qualify. Grants may be made to individuals or groups (including BCRA Special Interest Groups), who need not be employed in universities or research establishments. Information about the Fund and application forms Research Awards are available from The BCRA Administrator (address at foot of page).

GHAR PARAU FOUNDATION EXPEDITION AWARDS

An award, or awards, with a minimum of around £1000 available annually, to overseas caving expeditions originating from within the United Kingdom. Grants are normally given to those expeditions with an emphasis on a scientific approach and/or exploration in remote or little known areas. Application forms are available from the GPF Secretary, David Judson, Hurst Farm Barn, Cutler's Lane, Castlemorton Common, Malvern, Worcs., WR13 6LF. Closing date 1st February.

THE E.K.TRATMAN AWARD

An annual award, currently £50, made for the most stimulating contribution towards speleological literature published within the United Kingdom during the past 12 months. Suggestions are always welcome to members of the GPF Awards Committee, or its Secretary, David Judson, not later than 1st February each year.

BRITISH CAVE RESEARCH ASSOCIATION PUBLICATIONS

CAVE & KARST SCIENCE - published three times annually, a scientific journal comprising original research papers, reviews and discussion forum, on all aspects of speleological investigation, geology and geomorphology related to karst and caves, archaeology, biospeleology, exploration and expedition reports.
Editors: Dr. D.J. Lowe, c/o British Geological Survey, Keyworth, Notts., NG12 5GG and Professor J. Gunn, Limestone Research Group, Dept. of Geographical and Environmental Sciences, University of Huddersfield, Huddersfield HD1 3DH.

CAVES AND CAVING - quarterly news magazine of current events in caving, with brief reports or latest explorations and expeditions, news of new techniques and equipment, Association personalia etc.
Editor: Hugh St Lawrence, 5 Mayfield Rd., Bentham, Lancaster, LA2 7LP.

CAVE STUDIES SERIES - occasional series of booklets on various speleological or karst subjects.
No. 1 *Caves & Karst of the Yorkshire Dales;* by Tony Waltham and Martin Davies, 1987. Reprinted 1991.
No. 2 *An Introduction to Cave Surveying;* by Bryan Ellis, 1988. Reprinted 1993.
No. 3 *Caves & Karst of the Peak District;* by Trevor Ford and John Gunn, 1990. Reprinted with corrections 1992.
No. 4 *An Introduction to Cave Photography;* by Sheena Stoddard, 1994.
No. 5 *An Introduction to British Limestone Karst Environments;* edited by John Gunn, 1994.
No. 6 *A Dictionary of Karst and Caves*; compiled by Dave Lowe and Tony Waltham, 1995.

SPELEOHISTORY SERIES - an occasional series.
No. 1 The Ease Gill System-Forty Years of Exploration; by Jim Eyre, 1989.

CURRENT TITLES IN SPELEOLOGY - from 1994 this publication has been incorporated into the international journal Bulletin Bibliographique Speleologique/Speleological Abstracts; copies of which are available through BCRA.

BCRA SPECIAL INTEREST GROUPS

SPECIAL INTEREST GROUPS are organised groups within the BCRA that issue their own publications and hold symposia, field meetings etc.
Cave Radio and Electronics Group promotes the theoretical and practical study of cave radio and the uses of electronics in cave-related projects. The Group publishes a quarterly *technical journal* (c.32pp A4) and organises twice-yearly field meetings. Occasional publications include the *Bibliography of Underground Communications* (2nd edition, 36pp A4).

Explosives Users' Group provides information to cavers using explosives for cave exploration and rescue, and liaises with relevant authorities. The Group produces a regular newsletter and organises field meetings. Occasional publications include a *Bibliography* and *Guide to Regulations* etc.

Hydrology Group organises meetings around the country for the demonstration and discussion of water-tracing techniques, and organises programmes of tracer insertion, sampling, monitoring and so on. The group publishes an occasional newsletter.

Underground Photographer Magazine. This magazine was first published in December 1995, 48pp A4 with black and white photos. Subsequent editions have colour photos and articles on cave photography topics.

Speleohistory Group publishes an occasional newsletter on matters related to historical records of caves; documentary, photographic, biographical and so on.

Cave Surveying Group is a forum for discussion of matters relating to cave surveying, including methods of data recording, data processing, survey standards, instruments, archiving policy etc. The Group publishes a quarterly newsletter, Compass Points (c.16pp A4), and organises seminars and field meetings.

Copies of publications, information about Special Interest Groups, the BCRA Research Fund application forms, etc. are obtainable from the BCRA Administrator: B M Ellis, 20 Woodland Avenue, Westonzoyland, Bridgwater, Somerset, TA7 OLQ.