A CARTIOMETRICALLY ORIENTATED GEOTECHNICAL DATABASE FOR THE BIRMINGHAM AREA

BY

JONATHAN VAUGHAN EARTHY

A THESIS SUBMITTED FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF ASTON IN BIRMINGHAM

FEBURARY 1983

THE UNIVERSITY OF ASTON IN BIRMINGHAM

JONATHAN VAUGHAN EARTHY

A CARTIOMETRICALLY ORIENTATED GEOTECHNICAL DATABASE FOR THE BIRMINGHAM AREA

A THESIS SUBMITTED FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

1983

SUMMARY

A study of the potential users of an urban geotechnical map reveals that they could all utilise the same, existing information - if it were capable of presentation in different forms. A suitable multiuser, multiparametric, three dimensional, engineering geological map can be generated, stored and manipulated on an interactive computer. In order to save storage space, improve the spatial distribution and minimise sampling errors the data (site investigation reports, general records, maps and machine readable information) are pre-processed onto a square grid which is chosen to match the data distribution of the mapped area.

After an analysis of the possible locations and facilities, the system is written for a Harris 500 minicomputer. The data are stored as numeric equivalent codes and integerised real variables packed onto hierarchical, magnetic disc, direct access files. Routines are written in FORTRAN 66 (using GINO-F and a raster control library). Output formats are maps, sections and single archetypal boreholes, based on a (minimum) resolution of the 100m square 'segment'. Vector and raster graphical techniques are developed for conversational displays of geological information. These displays may be shown on VDU screens, X-Y plotters and printers.

The datastore creation process is examined and recommendations are made for its improvement. A user's manual for the 'map' is included, as are detailed descriptions of the programs and datastructure. The use of the database in the controlled interpolation of a three dimensional model and the production of a set of new maps (including an enhanced rockhead surface and a hazard map) is described. The advantages, disadvantages and potential of this spatial information system are reviewed and the possibility of transfer to a microcomputer is discussed.

KEY WORDS

Geotechnical database, Engineering geological mapping, Interactive computer graphics, Geological spatial information system, West Midlands

ACKNOWLEDGEMENTS

The project was supervised by Dr J.A.Morton of the Geological Sciences department and funded by an Aston University research studentship.

During the course of the study a number of other departments have provided assistance. Special thanks should go to: B.Mazey and E.W.Garnett of the Structural Engineers' department and Dr G.Demidovich of the Maps and Plans department (both at Birmingham City Council), R.Harris of the Severn Trent Water Authority and finally A.J.Bell of Aston University Computer Centre.

I would also like to thank Dr P.D.Mallinson, S.M.Holt and R.H.Beswetherick for their guidance in the more abtruse regions of the use of the Computer Unit's peripherals and software.

CONTENTS

This set is bound separately from the main thesis.

LIST OF TABLES

LIST OF PLATES

LIST OF FIGURES

1 INTRODUCTION

The majority of site investigation data is wasted. After its employment for a specific development, the report on the ground conditions at a site is filed - and never used again. This is because the information is not in a format which can be readily applied to any other task.

The aim of this project is to use all the data available, together with reasonable corrections to, extrapolations from and interpolations of those data, to produce a document of the greatest usefulness to the community. Such a document, which will almost certainly be some form of map, will be designed to benefit a number of groups to whom prior knowledge of ground conditions in the mapped area would be an advantage.

Unfortunately, all the interested parties will wish to use the different types of data, which are collected in different ways, to solve different cartiometric problems. The advent of the digital computer has introduced the possibility of purely 'cartiometrically orientated' cartography. This means that spatial relationship and measurement problems may be solved directly, without recourse to a physical map, provided that the necessary specific data are available. This approach might be called no-map cartography (Yolei in Davis & McCullagh, 1975).

The geotechnical map, in fact, meets all of Briggs & Briggs' requirements (Merriam, 1969) for the employment of a computer: the volume of data is large, mathematical manipulation is required and retrieval of information from a large databank is necessary.

Degani (in Freeman & Peroni, 1979) describes a 'map' which is so complex that it requires a computer to aid in its display as an Automatic Spatial Information System. He defines such a system as: 'A dynamically updatable body of data, interfaced with a body of models programmed for computer transformation of these data into spatial information in order to satisfy specific needs of defined users within the framework of well defined concepts and technology'.

The construction and use of such a computer database map of Birmingham city centre is described herein. The treatment of errors and the processing of the data is considered in some detail, because in this type of information system the concise display of the reliability of the data is as important as the data themselves. The reasonable 'spreading' of existing information, to predict the ground conditions in areas not actually investigated, is studied. A first step in an interpolation of this type is taken, criticised and proposals for a more sophisticated method are mooted. As a final test of the map database system a set of real maps are produced. These maps are intended as a demonstration of the flexibility of the 'map', but they also provide a usable index of the ground conditions of the Birmingham area.

The project resolves itself into two levels, the production of a usable map from existing data by computer and an analysis of the development and potential of computerised maps in geology. This duality leads to a number of disparate areas of study. A new urban geotechnical map for Birmingham is produced from data not specifically collected for mapping. The reduction and processing of these data employs many of the techniques developed in the Tyne and Wear project (Dearman et al, 1973) on a new area. One such technique is the unusual, but very effective, step of averaging the captured information on to a regular grid.

The digital computer is employed extensively for checking, processing and displaying the information. Computerised geographical interpolation techniques are tested on geotechnical data. Interactive, colour, raster graphics are used in the computer-aided processing of geological information. Environmental sieve mapping is tested as a computerised analysis technique.

The establishment of a hierarchical structure of: the index maps, the database and the original information is intended to improve the accessibility of the filed site investigation reports. For, by providing an index to the original documents, it will facilitate a simple access to the increased resolution and depth of the site investigation information. Conversely, from the original data to the map database and then to the index map set the ease of understanding and the simplicity of use increases. This will, for the first time, make the extensive geotechnical information which is recorded for the centre of Birmingham available to everyone to whom it is of use.

1.1 Previous Work And Its Results

1.1.1 Geology of the area

The Midlands have been studied since the first days of geology. All the early papers were listed by Harrison (1894) who pioneered the synthesis of observations. He reported the presence of three glaciers in the area and proposed the name of Lake Bosworth for the source of the lacustrine deposits but failed to find any evidence for an interglacial period or for subdivision of the glacial deposits. Eastwood et al (1925) described the Birmingham drift as 'Triassic and Coal Measures overlain by great formless masses of sand and gravel flanked, to the northwest and southeast by clayey deposits'.

Subsequent investigation was summarised by Wills (1937) and Shotton (1953) who produced papers that synthesised observations on the glacial deposits of the area and attempted to establish the chronology of glacial advance and retreat in the Midlands. Duigan (1956) dated a peat deposit found during the construction of the power station at Nechells as Hoxnian Interglacial. Mapping projects of north Birmingham by Kelly (1964), southwest Birmingham by Pickering (1957) and central Birmingham by Markham (NA) detect a first glacial advance during the Anglian. This produced three levels of lacustrine sediments, Lake Harrison being the lowest, from lakes trapped against the Clent-Lickey hills and Bredon hill. The advancing glacier eventually overrode these deposits. Kelly (1964) adds that this advance restricted the occurrence of the Early deposits to infills in the preglacial land surface. He explains the Nechells deposits by the infilling of a small lake during an interglacial period and correlates them with deposits at Cardigan Street and Washwood Heath. The Upper series is dated as Wolstonian and represents a triple advance of valley glaciers into the area giving lacustrine, fluvio-glacial and other types of deposit. A final glacial advance covered the area and completed the infilling of the valleys.

Horton (1974,1975) added the new Midland Link motorways data to the existing surveys, to more closely define the 'deeply dissected' preglacial surface. He found four basic engineering types in the superficial deposits and proposed that the Nechells deposits were contemporaneous with the Quinton lake deposits.

In historical times the city of Birmingham developed to supply manufacturing industries, many of these are now in decline. This has generated large areas of redundant land which are now available for redevelopment (Birmingham county structure plan & Cantell, 1977).

1.1.2 Geotechnical mapping

The need for a standardised notation for the geotechnical maps and records resulting from detailed site investigations, constructional records, extended studies (such as landslip or seismic zones) and national studies has long been recognised. In 1957 the British Standards Institute published a standardised notation for site investigations. Fookes (1967) suggested techniques for the organisation of site investigations, including critical path analysis for planning surveys. Two Geological Society working parties produced a detailed series of recommendations for boreholes (1970) and maps (1972). These cover all aspects of static and dynamic mapping of geological and geotechnical parameters. Further to these recommendations UNESCO (1976) has published suggestions for standardised map scales, types and nomenclature. Policies and projects for the use of these standards are given in the following papers: Price (1971), Matula (1978), Dearman et al (1979) (the Tyne and Wear project). Lopez Prado and Pena Pinto (1979) summarise the problems of 'selling' the resulting data.

The theory of geotechnical mapping has been advanced through practical application. The various attempts are well documented, but relatively few in number. Because, especially in the United Kingdom and Europe, the map usually results from a commercial need rather than a purely academic study, only a few parameters are documented in detail. Reported research ranges from Rockaway's (1976) analysis of the influence of data distribution and scale on maps to Hutchinson's (1977) summary of the assessment of slope dynamics. A high proportion of the studies carried out are dynamic, looking at landslip and seismic effects as well as potential ground movement - due to subsidence etc.. These studies are of extended sites, such as road routes, Brunsden et al (1975) or large seismic areas, Bracinac et al (1978). Many medium or small scale urban planning and country-wide summary maps are being produced in both Europe and the USA. Most of these are based on the pioneering work of Matula (Dearman & Matula, 1976).

Classification and reporting of rock and soil types, with the aim of improving the standard of presentation of geological data, have been studied by Stewart et al (1976), Franklin et al (1971) and Glossop (1969). The determination of an overall index of 'rock quality' by the combination of several parameters of a sample has been attempted by Cottis et al (1971) and Hwong (1978).

Some notable projects, at various scales, which provide clear illustrations of the problems of mapping and site investigation are as follows: Ward et al (1968) assessed the site for a large scale proton accelerator at Mundford, a survey of an unusual pseudo-extended site; Bazley (1971) and Wilson (1972) produced an engineering geologist's map of Belfast which makes a good attempt to solve the problems of presenting three dimensions, and showing a large number of discrete parameters, on a one or two layer two dimensional map; Brunsden et al (1975) made the Taff Vale road a testing-ground for morphological mapping; the study of urban Tyne and Wear by Dearman et al (1977) gave experience in archive searching, geotechnical databanking methods and display maps for urban areas; Cratchley et al (1979), studying the Maplin airport site, used computer techniques on an extended area to cope with at-depth information. In Europe, Gounon (1979) and Vidal Font (1979) used existing data to produce planning maps of Nice and Toulouse aimed at the wide group of non-engineering geologists. Finally, Bennett (1979) presents a perfect site investigation - analysing the results obtained for errors and considering the history of the area he produces a secure site model, rather than the usual disclaimed table of bearing capacities.

1.1.3 Geological databases

In 1961 G.A.Young, in the discussion of Bouma and Nota's proposal for detailed graphic borehole logs, suggested that, if numeric codes were used instead of graphic symbols, the logs could be read and stored by a digital computer. The suggestion was criticised, but not ignored.

By 1967 the first databanks of geological data were being constructed (Dillon 1967, Buller 1967). The Canadian Geological Survey have taken a special interest in computerised data storage, forming COGEODATA (Brisbin & Ediger, 1967) which studies and reports on developments over the whole spectrum of 'computers in geology' (Hubaux,1973).

Loudon (1967) produced an early data processing package, ROCKDOC, and subsequent developments follow his patterns: Harbaugh and Merriam (1968), Davis (1973) and Loudon (1979). In the United Kingdom the Institute of Geological Sciences (IGS) piloted the development of databanks of boreholes, Gover et al (1971), and their use for output and analysis of data, Rhind and Sissons (1971). Varey (1977) gives a detailed account of the early years of these, and other databanks in his proposal of a geotechnical databank system for use on a batch computer.

Outside the specific field of geological databases a large amount of work, much of it military in genesis (Oil, Defence and Planning being the three largest spenders in this area), was summarised at the NATO spatial data processing conference in 1973, Davis and McCullagh (1975). In those days interest was centered around the capture, interpolation and display on paper of three dimensional data. However, Schmidt and Zafft of the Harvard Graduate Design School noted that: 'the resultant integration of man and machine can be a paradigm for other (computer) fields, since it weds the computer's ability to manipulate data with man's ability to seek out, recognise and interpret these patterns'.

The development of geological datastores continued. In 1977 Odell produced the conclusion of the Cambridge group's work on geological data description: a nomenclature system, LSD02, and a suitable datastructure called CGDS. This completed a period of work on archive and museum storage systems, Cutbill (1971). A commercial system for the databasing of geological data, G-EXEC, was announced by Jeffery and Gill (1977). Varey (1977) and Cripps (1979) completed geotechnical data storage and retrieval systems for urban environments. Meanwhile, Tuckwell and Sadegrove (1977) tested a microfilm databank for site investigation reports after voicing the opinion held by a number of individuals: that mainframe storage of many types of geological data was too expensive to be justified outside an academic environment.

Recent work in the use of geotechnical databases attempts to follow Hubaux (1973) in: 'efforts to obtain a clear, uncryptic presentation of (these) data are as important as good input systems'. Morin (1979) carefully defines the users of the final document and tries to present maps of the simplest form and highest value to specific individuals by reprocessing the information derived from a Canadian geotechnical databank. Thinking as an engineering geologist he attempts to be selective in approach but also make the best use of all the data available, even to the extent of modulating the reliability of the result. In doing so he turns aside from conventional computer-aided map making which selects a fixed, optimum relationship between terrain digitising and contour line regeneration to realise a specified precision.

Reekie et al (1979) in Newcastle recommend the averaging of data into a square mesh grid. This has many advantages: the supression of rogue results, reduction in databank size, easy display of and reference to data. Reekie's paper also suggests the encoding of man-made features and hazards, as well as great simplification of data prior to encoding - to make another reduction in the size of the datafiles.

Whilst updating the engineering geological map of Rouen Buisson _et al_ (1979) used a geotechnical database, FIDGI, and an interactive computer-aided borehole representation / map alteration program, VERCOURS, to interpret the bedrock surface, rather than the more conventional fully computerised interpolation techniques. Their conclusions show that, using computer-aided methods, a geologist can view and utilise a large amount of data with great ease, to achieve a much more realistic final synthesis than any purely computer method. In the same vein, the latest NATO spatial data conference, Freeman and Peroni (1979), is concerned with more usable datastructures, conversational interrogation of databanks and the use of the more modern display devices. It must be noted, however, that a disturbingly large proportion of current investigation is concerned with finding uses for almost unsolicited, but readily available, data (i.e. LANDSAT and other remotely-sensed images) rather than deciding which are the most useful data and devising suitable data capture techniques.

1.1.4 Surface processing techniques

Two techniques of interest to the current project are interpolation and potential surface analysis. The interpolation of three dimensional surfaces is a long-standing research problem in which methods are often defined by national bias. Three main types of technique have emerged, based on different requirements. Most of the early work was intimately associated with automatic contouring but nowadays the uses are more widespread, as indeed are surface display techniques. Potential surface analysis is a method of analysing spatial models by the scoring of grids of weighted variables.

1.1.4.1 interpolation by weighted averaging

The simplest and most efficient method, this technique is employed extensively in the production of digital terrain models (DTMs) for topographic maps. It is the basis of most commercial systems and is the necessary first stage in trend surface analysis. Shepard (1968) described the techniques needed in detail and developed the weighting factor, used by the Ordnance Survey (O.S.), to give linear biassing for dense data. The same algorithm has been published by Davis (1973) and was used by Varey (1977).

An enhancement (IBM,1965) to this distance-related weighting uses local slope values, calculated at each point, to improve the realism of the averaged values for each node of the regular grid. This method is used by the Experimental Cartography Unit (ECU) in one of the best interpolation routines currently available. However, when Coe and Cratchley (1979) experimented with this package they found that when used with sparse, real data the most accurate and 'natural' surfaces were obtained with the slope parameter minimised. Simple averaging has the advantage that the surface interpolated approaches a plane in areas of little data. This is distinctive, minimises errors and cannot produce values far above and below the existing datapoints - as happens when the following, more sophisticated, techniques 'blow up'.

1.1.4.2 curve fitting

This is one application of trend surface analysis and is used, mainly in the UK and USA, for filter smoothing of averaged maps and the interpretation of map and surficial structures of many spatially-related parameters.
Krumbein (1959) proposed the technique of three dimensional polynomial curve fitting to a DTM in order to analyse the properties of the surface. Obviously the fast digital computer is invaluable in this work, allowing a large number of surfaces to be 'tried out' against the original whilst displaying the difference, or 'residual', in graphical and/or numerical form. Since the advent of the Fast Fourier transform harmonic analyses can be performed cheaply. These are of greater use than polynomials to the geologist, allowing the interpretation of harmonic features such as sedimentary dynamics and fold systems. This work is summarised by both Whitten and Robinson (frequency analysis) in Davis and McCullagh (1975). The recent interest in the fitting of regionalised functions, the 'spline-fitting' techniques, which produce very realistic surfaces is the subject of a paper by Werner in Freeman and Peroni (1979).

1.1.4.3 kriging

Employed mainly in Europe and South Africa, this is another sophisticated regionalised technique, using pseudo-statistical moving averages to minimise the error in a regionalised variable. The process simultaneously produces a surface and a 'variogram' of estimated error. This byproduct is useful in estimating the type and accuracy of the calculated surface. It was designed for, and finds its main use in, mineral resource prediction, although its devotees use it for all spatial interpolation, Buisson <u>et al</u> (1979).

1.1.4.4 potential surface analysis

This is a method of sieve analysis to combine sets of spatially related integer and real variables to achieve an overall value for the 'potential' of the elements of an area. The foundations of the process were proposed by Zetter (1974) for rural planning. Ammer et al (1976) laid down guidelines for its use and Pocock (1979) developed his advanced 'Typical Area Element' techniques to 'systematically assess the potential of an area to accommodate a particular type of development or land use'. 'TAE' provides an interpretative matrix for the scores as well as the potential map itself.

1.1.5 Computer graphics

This science commenced in 1953 with the Whirlwind computer, to display output for the United States defence system. The first interactive graphics were developed by Sutherland for the TX series of machines in 1955. By 1961 dedicated graphics computers allowed commercial, but rather expensive, light pen interaction. One of the first computer-aided mapping projects produced the Oxford atlas of Britain (1963). This was a conventional publication made easier to produce by the computerised storage of data. Throughout the Sixties the strengths and weaknesses of the computer in picture display and graphic data processing were studied, forming the basis of most of the mapping and graphics packages available today. The enormous cost of the early mainframe computers and graphics peripherals lead to the development of schools of study, in which this pioneering work was carried out. Of interest to this project are: the Harvard Graduate Design School (SYMAP by Fisher, 1968), the

Computer Aided Design Centre (GINO graphics package, 1976), the Experimental Cartography Unit (map making and spatial data processing, 1969) and the Utah Department of Engineering Graduate School who, in the early Seventies, produced most of the fundamental algorithms for realistic three dimensional graphics (Newman & Sproull, 1981).

The practical use of mapping programs, particularly the SYMAP package, is demonstrated by Rosing & Wood (1971) and Taylor et al (1976) in atlases of Birmingham and Tyne and Wear. Leclere (pers. comm.) used potential surface analysis to produce a manually created map of hydrogeological parameters intended for waste disposal pollution control in the Midlands. More sophisticated forms of output are discussed in Davis and McCullagh (1975) and Freeman and Peroni (1979). However, it should be noted that even if a program is being used sucessfully by one unit or individual there is no well established means of distributing that work to other individuals and indeed, the language and devices used by that program may not be reproducible at any other site. Thus a large amount of conscious "re-invention of the wheel" takes place. Most of the information is transferred verbally or by simple published schemes (for the solution of specific problems) from the few international conferences.

Recent advances in computer graphics centre around the availability of cheap computer power in colour raster graphics and microcomputers, new forms of presentation (such as poster sized flat displays and 'real' three dimensional displays) and the invention of 'fractals'. These are an application of fuzzy theory by which the computer is made to 'imagine' the details in a geographic scene. The appropriate resolution for a particular viewing scale is created mathematically. The future for computer graphics promises to be even more exciting, with cheap image processors for remotely sensed data analysis (Harrison, pers. comm.) and aircraft flight simulator quality graphics on home microcomputers...

2 DEMANDS AND REQUIREMENTS OF THE MAP

2.1 Potential Users

The opinions of a number of possible users of a 'geotechnical map' of unspecified but potentially very flexible format were sought in informal interviews. Due to a rather alarming unfamilarity with the concept of planning maps, an element of prompting was introduced into the questions for all groups.This resulted in an increase of interest in most groups. The general conclusion was that such a map would find use, if the quality of the data was guaranteed. As with computer software exchange, interest waned rapidly at the mention of payment for access to this 'map' although such a commercial approach is often used in the professional sphere to imply an occasionally erroneous greater reliability in the data for sale.

2.1.1 Interviews

Two groups within Birmingham City Council were contacted. The Planning department, who deal in large budgets, were obviously worried about the credibility of derived data but found the concept of a large scale costing or a relative development difficulty map for the whole Birmingham area of greater interest than the more detailed products. The Structural Engineers' department deal with the maintenance of existing constructions and already bank site investigation records. They decided that the display of detailed information on a small site was of more relevance to them for the estimation of specific site conditions, in costing small projects and predicting the type of equipment required.

The engineering geologist interviewed stated that he would want to see only original site investigation reports or order a new survey, whilst a site investigation contractor suspected that his drilling teams would prefer to copy results from the filed information rather than create new records. However, both groups showed considerable interest in the proposals. This was caused by an acute awareness of the amount of 'hidden' data which are collected but not re-used. They added that re-use is not generally practical because of the time required to extract information from the site investigation report format. The facility to quickly show the data in a number of ways (plans, sections etc.) was considered to be one of the most useful aspects of the study.

Architects, like the planners, are interested in maps but would prefer to have a number of discrete parameters available to make their own judgments for a specific requirement. But, the use of a combined 'risk' map was also of interest - if the factors used were clearly listed. Architects also appreciated the use of more sophisticated forms of graphic presentation, such as the isometric plots, to enhance certain features of a map.

2.1.2 Conclusions

A simple conclusion from this survey would seem to be that none of the above groups want the same thing. But, in fact, all of them can use the same data, if it is presented in different forms. Other points are the general requirement for 'worst case' information and a need for an estimate of the reliability of these data. An interest in an estimate of development or survey cost, or a general hazard index is also common, though expressed in different ways. Scientific use alone has different requirements and it is possible that research users are the only group who will not be adequately supported by the results of the current project.

2.2 Choice Of The Area To Be Mapped

2.2.1 The existing data

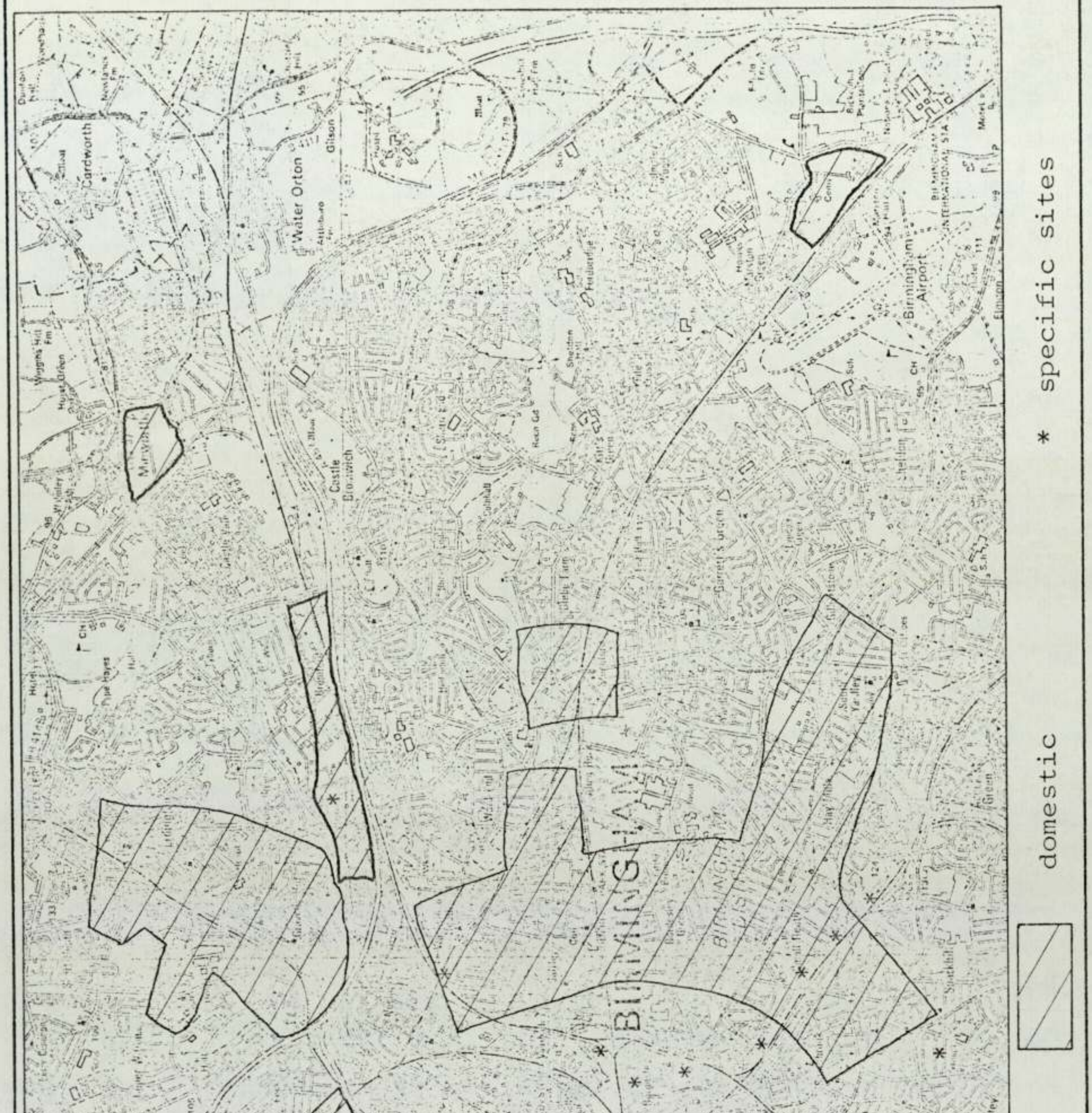
This study is in many ways a descendant of Varey's 1973 project on borehole encoding and as such it has inherited 524 fully encoded boreholes, held in machine readable form at Aston. In the planning stages of the project it was felt that this arduously acquired data should be re-used. Thus, the area chosen for mapping had to be contiguous with Varey's which, because of its naturally high density of site investigation reports, is Birmingham city centre.

2.2.2 External requirements

Another important feature of the project is to provide a document of use to the construction community in the selected area. Consultation with the city planners revealed that the development boom of the early Seventies had finished, leaving only a few developments in the public arena. Private industrial development was being encouraged at Minworth, see Figure 2.1, and along the M6 between Gravelly Hill and Castle Vale with some scattered 'soft' redevelopment land still to be exploited within the middle ring road.

Housing land and developments are much more widespread, specifically in: Winson Green, Handsworth, Birchfield, around Witton station, Erdington, Gravelly Hill, Stetchford, Saltley and along the Alum Rock road, Small Heath, Sparkbrook,

Figure 2.1 Map of Development Areas in Birmingham



Hay Mills and South Yardley. It should be noted that a large amount of this work will be redevelopment rather than new building but nevertheless extra information on the ground conditions will be required.

2.2.3 Availability of information

Because new, large area site investigations are rather expensive to commission, one of the principal sources of data for a store of geotechnical information is the old site investigation reports. It is important to choose an area which has not only a large number of reports on it but also a reasonably even spread of these reports over that area. Luckily, Birmingham has had new road and sewer systems as well as the recent spate of large buildings and in most, if not all, parts of the city and its village satellites some reports of rather variable quality are available.

2.2.4 Physical limitations

A significant problem with a limited - manpower, limited - time project is the time available for the capture of the necessary data. This period cannot always be judged at the commencement of encoding and some flexibility should be allowed. If the time allocated proves insufficient, information which has been collected with a freedom in the area to be mapped will be of a higher quality than on an oversized fixed area, collected with a freedom in the number of points to be taken.

A final point is the convenient size of the maps produced. The scales of suitable base maps are fixed and therefore the dimensions of the area mapped should be selected, with due regard to more important considerations, to make use of as few base maps as possible and be of a physical size which can be easily manipulated.

2.2.5 Location of the area

Bearing all these considerations in mind, the area to be mapped, see Figure 2.2, was chosen to be as follows: From Edgbaston (SP0684) in the southwest to Perry Barr (SP0692) in the northwest and extending east between these limits as far as coding time allows to a minimum eastern edge set by the eastern extremity of Varey's data near Ward End (SP1284,SP1292). In fact, only this minimum area was mapped, to allow sufficient time to be spent on data interpretation. The area includes urban, suburban and redevelopment areas and exhibits a large variation in site investigation report density. This will test the data capture and display techniques as much as possible. Because the whole mapped area lies in the Ordnance Survey SP coordinate block only the numeric coordinates are used in the remainder of this document.

2.3 Description Of The Selected Area

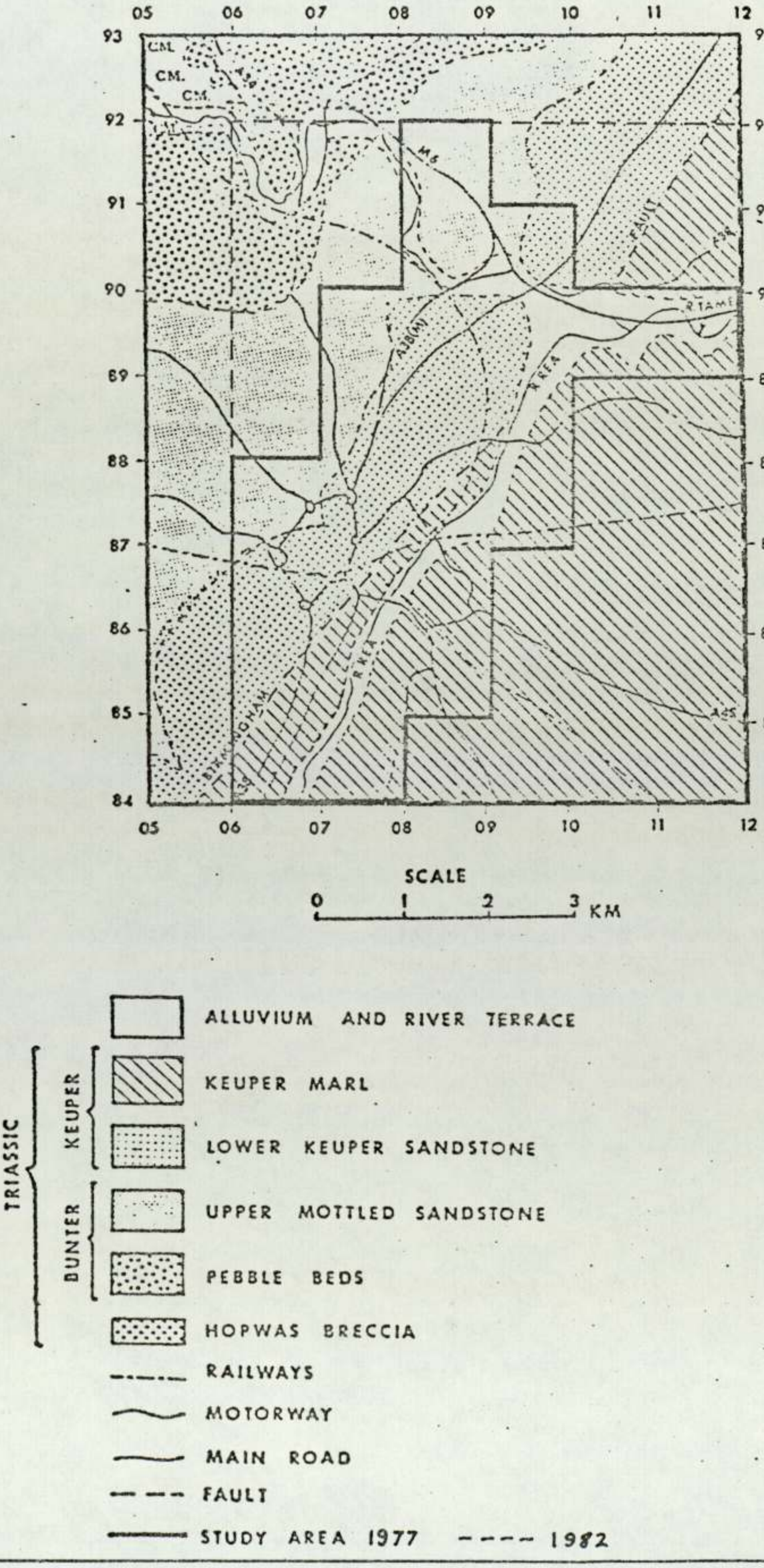
2.3.1 History

The city of Birmingham has a relatively short history. It has grown since the beginning of the seventeenth century, from a small iron-working village with a population of three thousand, through a rich manufacturing town in 1700, to 'the Toyshop of Europe' by the beginning of the nineteenth century. Even then it was "not a place a Gentleman would choose to make a residence", due to the overpopulation, noise and pollution.

From 1850 the surrounding villages expanded, giving the area a population of 300,000 in 1861. By 1914 Birmingham city was judged to contain some of the worst housing

Figure 2.2

GENERAL GEOLOGY OF BIRMINGHAM AND STUDY AREA



conditions in the United Kingdom. In 1947 the population of the conurbation reached 2,240,000, over one million of these resident in the city itself. As a result of this spectacular, unplanned growth and subsequent collapse of intensive manufacturing industry in the area many industrial features and man-made deposits have become a part of the soil of the city. These brick pits, sludge lagoons, flowing coal tar and petroleum residues and toxic waste burial sites are as important to a geotechnical study as the natural geological features.

2.3.2 Present day

The present day city area under review (see the base map in Appendix 7) includes the city centre with its new high-rise blocks, shopping centres and urban motorways. This construction is very recent and, in fact, provides the majority of the data for this map. To the east of the city is a band of industrial and development land made up of a complex mass of communication networks (canal, road, rail & motorway), warehousing, manufacturing property and, often derelict, utilities. This zone spreads out along the railway lines into the inner suburbs of Victorian and pre-war housing. There are large tips at Adderley and Sparkbrook. In addition to the parks, some areas of open ground occur in Witton.

Naturally, high concentrations of good data are only available for areas in which development has been completed and <u>vice versa</u> for the new areas for which site investigations are being planned.

2.3.3 Geology and hydrogeology

The geological and geotechnical properties of the sediments underlying this area are summarised in Table 2.1. For a more detailed description and analysis the reader is referred to Horton (1974,1975) and Chandler et al (1968). The bedrock dips gently to the southeast and the most abundant deposit is the Keuper Marl. The main structural feature is the Birmingham Fault. This fault still causes some differential settlement problems in buildings located over the fault zone.

The hydrogeology of the area is complicated by the effects of the old and damaged drainage system and the existance of many perched water tables in the clay and sandstone horizons of the Glacial drift and Triassic sandstones. The permanent water table is deeper than 100m. The Bunter aquifer provides a part of the water requirement of the city and care has to be exercised in preventing contamination of this supply.

2.4 Conclusions From Previous Work

2.4.1 Geotechnical maps

Lopez Prado and Pena Pinto (1979) laid out the function and purpose of urban geotechnical maps on a medium scale as:

DEFINITION of geotechnical and land divisions

ANALYSIS to give fundamental characteristics

ASSESSment of these values geotechnically and constructively

The information displayed should include the internal and external dynamics (slumping, slipping), substrate and superficial geology, hydrogeology, relief, natural risks, environment and relevant derived geotechnical parameters.

Table 2.1 Geology and Technical properties of the Sediments

	TYPE	OUTCROP	THICK-NESS	LITHOLOGY	COLOUR	GEOTECHNICAL	HYDRO-GEOLOGICAL	NOTES
P.G. & RECENT	fill	general	30m Adderley	common ash, highly variable	–	local study required	variable	cellars and industrial waste common
	post glacial	Tame, Rea and Cole valleys	10m lower Rea	fine grained occasional organics with coarse material in the lower units	black, grey or brown	variable, organics can cause differential settlement	coarser units may cause problems	often re-worked by man
PLEISTOCENE	Boulder Clay and unbedded	most of the NW half and the Washwood Heath and Small Heath areas	total 35m Aston Six-ways	gravelly clay plus rock fragments	greyish or reddish brown	generally adequate for foundations	usually impermeable	strongest fine grained drift deposit
	Glacial Lakes	interbedded and scattered		variable, bedded, fine silts and sands with common clays and occasional sands and gravel	reddish or yellowish brown	variable strengths rather plastic	complicated and unstable	individual survey required
	Glacial sand and gravel	Witton, Wood End road and Hay Mills		well sorted sands and gravels, some clayey silts	reddish brown matrix	loose to dense good foundation	dewatering may be required	provides fill material
	Inter-glacial	Nechells	(10m Vauxhall)	peaty, laminated clays, silts, clayey silts	dark and greyish brown	plastic, compressible and often very wet	low permeability	individual survey required
KEUPER	Keuper Marl	south and east of city, Erdington and Gravelly Hill		mudstone with bands of silt and occasional sandstone	dark red or reddish brown with green bands	the weathered rock often becomes incompetent as a foundation	siltstone bands (skerries) increase the permeability	high concentrations of sulphate are common
	Lower Keuper Sandstone	band from city centre to Stockland Green		poorly cemented sandstone, pebbley or with interbedded mudstone	yellow-buff or red-brown	excellent foundation	water perches due to clay bands	
BUNTER	Upper Mottled Sandstone	band parallel to the Lower Keuper Sandstone through Aston park		fine weakly cemented slightly clayey sandstone with bands of silt or clay-silt	red often micaceous	similar to the Lower Keuper Sandstone	aquifer but contains clay bands	occasional 'flowing sand'
	Pebble Beds	west Witton		coarse false-bedded sandstone with well rounded pebbles that can reach 9" diameter	brown-red	good foundation	aquifer	only proved in a few boreholes

They found that the difficulties experienced in the use of geotechnical maps stemmed from novelty and lack of experience (unknown by planners, designers & builders), difficulty in interpretation for non-experts and lack of contact between planners and geotechnical experts.

Bickmore in Davis and McCullagh (1975), musing on the computer-aided production of the Oxford 1963 Atlas of Britain, highlighted the importance of accurate registration of datasets within the coordinate system employed in the construction of maps. He mentioned that a computerised map-making system would probably be less thought-provoking than a set of maximum resolution maps, thus illustrating the common "numbness of mind" experienced when faced with a computer. He ends with the 'attractive, if theoretical possibility of interrogating the original data at any point, at speed, and hopefully at reasonable cost', giving rise to the concept of "Instant Maps".

Wilson (1972) threatened that 'Engineering Geologists, with a minimal grasp of civil and structural engineering should beware of a technical backlash which may produce geological engineers at least as competent to deal with the marginal territory of urban geology', pointing out that the geologist must understand that the engineer bears a professional responsibility for his work which is absent in pure geology and that geology in urban areas is much more mechanistic in type and requirement than its open - field counterpart.

2.4.2 Computers in geology

On the computerising of geology Hubaux (1973) noted the conviction of many geologists that 'Geological data, for some mysterious reasons, would not fit within the strictly ruled realm of the computer.', due to the subtlety of geological analysis. These fears can be resolved into the following real concerns: the enforcement of unsuitable measuring techniques which are a barrier to progress; encroachment on liberty of presentation; enslavement of the geologist to data collection; introduction of the 'drastically rigorous' method of mathematics; the simple fact that data coverage is just not good enough for use by a computer, that most data - intensive of analytical tools.

Hubaux (1973) adds that 'the first automobiles were designed to look as much as possible like horse-drawn carriages. Most of today's computer-based geological files are still conceived to mimic manual files'. Reekie's (1979) attempt to improve this situation, as programmed by K.Bolton at Sunderland Polytechnic, results in a very simple data store (based on card formats) which contains summaries of stratigraphy, lithology, moisture content, bulk density and Atterberg limits for up to eight layers and codes for man made features. The objectives of the databank are to:

reduce and standardise the data used
allow rapid retrieval of a data summary in map form
allow regular updating
facilitate pilot map compilation
force regularising of data distribution

In this project the data were averaged onto a 100m square grid because:

it was related to the National Grid
there is a practical limit to the amount of data which can be shown on a map
this unit area functions as a point on the maps used

the number of lithological units is not too large or small (this depends on the area and its geology)

the size of the area should significantly reduce the number of points

The resultant system is an 'instant map' which prompts for input to specify the information to be presented on a 1km square typed map. This is a great advance from making up one's own map but its rather limited function illustrates a comment made by Degani in Freman and Peroni (1979), to wit: 'a computer specialist is a person who does not know what (spatial) data processing is all about, but can offer you a faster algorithm for it, compile it for the necessary data, classify these data faster and store them on less tape than you ever expected.' i.e. the programmer can only solve the problem he is given. This clearly suggests firstly that the greater the flexibility of the programs, the more hope there is of achieving the output you require, and secondly that it is not possible to get more out of programs or data than has been put into them in the first place. The careful definition of formats and functions at the beginning of a project is essential.

2.4.3 Programming

On the subject of programming a computer, the re-use of programs, routines or algorithms written by other people, with its potential saving in time and effort should be mentioned. The arguments for and against transfer, as summarised by Baxter in Freeman and Peroni (1979), are well rehearsed but totally contradictory.

For: prevents unnecessary re-invention, everyone gains, it is cheaper (in men, machines & time) than redevelopment, makes badly needed capabilities available to all, allows for a quick 'technological fix', costs are dependent on technical factors which are the major barriers to transfer.

Against: in-house skills are enhanced, job satisfaction is lower, cost savings require particular conditions, not all the capabilities are needed, the number of applications available to and acceptable for local circumstances is small, human frailties are the main barriers to transfer.

The experience from the current project suggests that non-professional and/or machine dependent programs are difficult to re-use, let alone transfer. However, the use of routines supplied as libraries of functions for a main program is easy and, in fact, almost essential to the speedy construction of complex computer systems. Verbal or diagrammatic descriptions (ALGORITHMS) give clear, easily understood and usable insights into another individual's solutions for a problem and should be employed as often as possible, for educational content as well as convenience.

2.5 Proposal Of Computer System And Final Map

2.5.1 Requirements

An urban geotechnical map cannot replace a full site investigation, but it has a real use in pre-site investigation planning. Additionally, an easily understandable geotechnical map can be employed to advantage in the initial feasibility and costing stages of urban project planning. If the ultimate use of such a map is born in mind when considering its design and requirements several factors become obvious:

The data on the final map should be simply presented, or interpreted for, the user.

It is of advantage to the user to quickly manipulate the data, in a way not possible with the existing site investigation format, to produce derived parameters of interest to that particular individual.

For a scientific user, or advanced mapping, spatial analysis and statistical interpretation techniques would be of great use.

The display of data and whether they are original or derived should be as clear and concise as possible.

Although the input data are poorly distributed the final map should give as complete as possible output coverage of the area as recommended by Vidal Font (1979).

With respect to the previous point, and the variable quality of the input data, measures of the accuracy and reliability of the presented information are just as important as the data themselves.

Because the area is urban, new work will usually be carried out on reclaimed land. Therefore a record of previous industrial activity and its remnants should be available, side by side with the more usual geotechnical parameters.

2.5.2 Data preparation

All of the potential data sources provide the greatest detail and the highest resolution. Following the aforementioned suggestions of Reekie (1979), Morin (1979), et al it is proposed to simplify the available information to provide a synthesised set of geological and geotechnical parameters - arranged on a regular mesh grid, to interface with the National Grid. This will reduce the problems of registering the datasets obtained from diverse sources on different scales. The averaging of data should also decrease the effect of errors and, along with some preselection of parameters to reduce redundancy, will greatly decrease the filestore requirements of the final datastore. The resulting records will be held on a magnetic tape or disc datafile for speed, ease of access and simple manipulation.

The size of the grid mesh has to be considered carefully. The statistical study carried out, Appendix 1, on Varey's (1977) data distribution indicated that a grid size larger than 150m square is best for 'feeding' interpolation programs. A set of experiments was also carried out on the highest point density area in Varey's databank. Using an interpolation program the 100m mesh was considered to be the best all round compromise for the various data distributions tested. The Nvyquist frequency, the minimum wavelength of feature which may be resolved, for this mesh is 200m. It has been shown that, in reality 250m is closer than 200m to the true minimum resolution (Davis & McCullagh, 1975). This value seems rather large but examination of the spatial relationships of the study area indicates that much larger features could be completely lost and therefore the occasionally high resolution available for the concentrated site investigations is, in terms of the whole map, anomalous.

The databank so far proposed will contain information on any 100m square which has a borehole located in it. To extend the 'map' to cover the entire area of study a computer - based interpolation system which simultaneously provides an indication of the proximity of hard data is required. If the program has to run every time an estimated point or a whole borehole is required the response time of the system will be very poor. Therefore all the data should be interpolated once and stored on the same grid structure as the original data. Subsequent operations will require only a 'read data and display' operation rather than a far longer 'find a set of nearest points, interpolate for each parameter and display' program.

Updating of records or display of data is best carried out using an interactive program suite which will interface with raw encoded data or the interpolated geological model.

2.5.3 The soft computer map

It has already been noted that even the best interpolation programs have some problems, Coe and Cratchley (1979). To program a computer to carry out a full geological interpretation would probably be impossible. The VERCOURS program, Buisson et al (1979), allowed manual extrapolation from borehole information but some difficulty was experienced with non-linear datapoint positions. Averaging and interpolating the records on to a grid circumvents this difficulty and reduces the display scaling problems. It is intended that the interpolated data provide a starting point for input to an interactive geological mapping/extrapolation program for the interpretation of appropriate datasets, using advanced interactive graphics displays for input and output.

The final model of geological and geotechnical parameters can be accessed by computer to provide single segment records and parametric maps of the highest quality. This system will allow the simple representation of many discrete or linked datasets, both on a visual display unit (VDU) and permanently, thus solving the problems of confusion and inaccuracy experienced when reading complex multiparametric maps by allowing the user to design his own from the source data.

2.5.4 The hard paper maps

For the planner, real maps are much more useful than a pure cartiometric database. In urban areas the scales of 1:10,000 and 1:25,000 would seem to be the most appropriate to use, UNESCO (1976). This implies that the minimum resolution on these maps will be on 1·0 and 0·4 cm grids respectively. Maps have been plotted on these scales and, for the uses specified, higher definition is not considered necessary. The final real map will be made up of several sheets considered to be of use in themselves and also to advertise the more sophisticated use of the computer map database.

3 ENVIRONMENT AND CONSTRAINTS OF A COMPUTER BASED MAP

This chapter summarises the state of the art in computing - indicating the uses, speed and relative costs of some of the more readily available devices in use today. Special reference is made to the facilities, installed at Aston during the course of the project, which have been employed or were considered for use in the making of this storage, retrieval and display system for spatially - related descriptive data.

3.1 Batch Processing Machines

BATCH PROCESSING is the serial running of each JOB submitted to the computer from a QUEUE of tasks. It is the result of a need for many individuals to use the same high cost facility and was the first system architecture employed for large computers.

These machines have an extensive capability to store data held both ONLINE, directly readable and accessible, and OFFLINE, held on magnetic media needing to be physically placed in the machine. This makes manipulation and storage of large sets of data very easy. The size, and nowadays the speed, of batch computers means that large and long PROGRAMS present no problem to the system. They can be run at a time convenient to the majority of users and the whole CORE is accessible to any program. It should be noted that, even though the mean time between failures is large on modern computers, long programs run the risk of a CRASH during their execution. Therefore it is standard practice to save significant values at regular intervals during a run. This

DUMPING means that a program can be restarted from the last DUMP rather than the beginning. Despite the many diagnostic aids, developing a program on a batch computer is a long job because of the slow TURNAROUND times. This tends to inhibit the fine tuning of programs for efficiency - leading to slow but safe routines. The USER pays no real penalty for this inefficiency, he does not wait at a terminal for an operation which might be performed faster.

In general, batch computers have greater depth to the SOFTWARE available. Several input and output options, BACKING STORE on many discs or tape units and larger collections of PACKAGES and LIBRARIES can be maintained. In addition sites with large computers also support many input and output PERIPHERALS. As well as line printers, graph plotters and terminals - due to the high total investment in the facility: tape and card readers, fiche printers and offline devices such as film recorders, digitisers, key to disc loggers etc. are available, allowing great freedom of choice to the potential user.

3.1.1 Examples

Two machines are available at Aston. One is an example of the earlier type of batch machine, an ICL 1904s which will be in use until mid-1984. It is old, slow with an increasing failure rate, indicating incipient senescence, but the faults are known and the OPERATING SYSTEM, a resident program which changes the box of components into a functioning computer, has been tested and tailored to the requirements of the site. The second machine, a CDC 7600 is based at the Regional Computing Centre in Manchester. The service offered is professional and is secure for the foreseeable future because of the national basis of the regional computer structure. The centre is an example of the solution to the continuing need for the enormous computing power and storage capacity of large batch machines. It offers a very fast computer system with modern mainframe features such as dual speed core and vector pipelining but not as yet the ability to parallel process, so useful in accelerating the analysis of spatial data.

3.2 Interactive Devices

The division between interactive and batch computers is becoming much less clearly defined as the power of small computers and the skills of operating system designers increase. However, the INTERACTIVE machine has more limited FILESTORE and features with fewer program packages. It also seems to generate a less professional attitude to the service provided among the associated staff.

Because the user is directly connected to his program in the core via a fast-responding TERMINAL the effective access to the machine is very high, thus eliminating the problems associated with turnaround time experienced in batch mode. This changes the character of program development, allowing the user to create programs very quickly and test improvements as they occur to him. Of course there is a danger of wasting time by following one small error in an expensive fashion or over-elaboration of programs but the advantages in producing fast, conversational programs that can be interacted with in verbal or graphical form are enormous. The style and efficiency of programs is also improved, because the performance directly affects the user.

The response time of a computer to a terminal user depends on the LINE SPEED, the rate of transfer of information to and from the VDU or teletype, and the number of other users with a share of the core time. Unlike waiting for output from a batch program the time spent waiting at the terminal is completely wasted. Extensive <u>in core</u> manipulation or TIMESHARE delays may lead to run times longer than are usually experienced with batch programs and, with the added factor of operator error, the risk of a crash is as real, as costly and even more frustrating than on a batch machine. For these reasons the FAIL-SAFEing or dumping of serious interactive programs should be considered.

3.2.1 Examples

Three interactive computers are accessible at Aston. An obsolete but very reliable Hewlett Packard 2000, a rather small machine with very few input/output (I/O) devices and little filestore. The Midland Universities Joint Computer: a DEC 20/60 based at Birmingham University. This has advanced graphics and input facilities at Birmingham. It is very FRIENDLY, conversational and easy to use i.e. a well written operating system, and reliable but had only limited filestore facilities at the start of this project and was slow and difficult to access from Aston. The third machine is a new and very fast Harris 500. This is accessible, being on site, via fast lines but is unreliable and not very user-friendly. The filestore available is reasonably large but no archive tape storage is provided.

Virtual memory allows the Harris to have a large effective core size, making it some six times more powerful than the ICL 1904s. It should be noted that the ICL has some interactive capability. This gives the user the ability to manipulate files but not to converse with programs. At risk of causing further confusion, the Harris has a 'background queue' for batch jobs and is, in fact, the best machine on which to develop and run very long programs, if fast turnaround is not important.

3.2.2 Microcomputers

The microcomputer is a new addition to the interactive computer field. Generally single-user it has the advantages of being small enough to be portable, cheap enough to be purchased from petty cash and very reliable. The MONITOR, or operating system, is usually friendly with great flexibility. This allows the custom building of a 'tool' for a particular operation and, of course, the cost is low enough to justify this approach.

Disadvantages associated with the micro centre on the slow speed, ½ to 4 MHz, and limited power of the current machines. However, the technical advances in microprocessor technology are incredible and the only problem in the future will be the provision of mainframe standard software for these devices.

3.3 Languages, Libraries, Storage And Peripherals

3.3.1 Programs

Computer programs are presented to the machine in the form of a series of logical or mathematical statements written in a language which can be read by a program in the computer, called a COMPILER, which translates the English-like instructions into machine commands. A number of languages exist to satisfy the requirements of different users. Table 3.1 compares the features of the more common languages available at Aston.

Table 3.1 Comparison of Programming Languages

NAME	VOCABULARY & FUNCTIONS	MATHEMATICS PROGRAMMING	LOGIC STRUCTURE	FILESTORE HANDLING	USABILITY EASE/SPEED	AVAILABILITY
BASIC	small	fair	poor	poor	fair(simple)	high, the language is often extended
PASCAL	small	fair	good	fair	good	fair and improving
FORTRAN IV	large	good	poor	fair	fair(common)	high
FORTRAN 77	large	good	fair	fair	fair	improving
ALGOL 60	fair	fair	good	fair	good	moderate
ALGOL 68	good	fair	good	fair	good	poor
COBOL	fair	poor	verbal	good	poor	fair-business
ASSEMBLER & machine code	fair	fair	complex	complex	poor	universal and is much faster in use

A number of the functions in a program are usually found to be logically discrete entities. These can be written as SUBROUTINES. Subroutines are small program segments carrying out an operation for the main program. If programs are written as a series of subroutines the development time and logical structure usually improve, although there is a possibility of loss of efficiency. If a number of subroutines can find use in several programs most computer systems allow the user to build up a LIBRARY of these useful program segments. They can be integrated into the required programs at CONSOLIDATION time, when the compiled program is translated into a usable CORE IMAGE. This saves the user the effort of re-typing the subroutines, leading to smaller programs and yet faster development.

3.3.2 Libraries

Libraries of subroutines are commercially available for certain specific fields of computing. They are well designed, usually efficient and allow the programmer to use the expertise of other programmers in his work without hunting for and re-coding algorithms. System libraries are attached in the same way as a user's libraries. Of interest to this particular project are mathematical, NAG, and graphics, GINO etc., libraries. These are available on most of the computers mentioned. Very little work has been carried out on the provision of libraries, or similar functions, for microcomputers.

3.3.3 Storage

Storage of the information in and around the digital computer can be carried out in a number of ways. These are summarised in Table 3.2. As with programming languages the

Table 3.2 Comparison of Filestore Types

OPERATION		DIRECT ACCESS	SERIAL ACCESS	MAGNETIC TAPE	PUNCHED CARDS	MICROFILM	DISKETTE	PAPER LISTINGS
CREATION		requires a user's program	system program or direct input	special system program	typing	system/ camera	system/ program	system
UPDATING		insert record	recopy whole file	recopy tape	insert card	no	as DA or serial	no
SIZE		design dependent	fixed but large disc size gives a large file	compact and very large	poor, high physical weight	most compact of all	variable but limited	bulky, not robust, flammable
ACCESS	batch	yes	yes	yes	yes	—	not necessary	—
	inter-active	good	yes	probably not	no	—	as DA & serial	—
	no computer	no	no	transferable between machines	yes	reader required	microcomp. is always accessable	good
LONG TERM STORAGE		no	no	yes	yes (bulky)	yes	yes (cheap)	yes (bulky)

Table 3.3 Comparison of Common Input/Output Peripherals

TYPE	INPUT OR IN- and OUTPUT	INFORMATION DENSITY	ALPHA-NUMERIC	GRAPHICAL	FORMAT	COLOUR	SPEED	COST
CARD PUNCH & READER	in	low	yes	no	80 column card	—	whole process is slow	generally available but machines are expensive
VDU (teletype terminal)	in & out	low	yes	no	video 80x24 characters	no	fast	relatively cheap
graphics VDU	in & out	fair	yes	yes	video free	possible	fast	fair to high
line printer	out	low	yes	no	paper 134 to 144 columns	no	fairly fast	generally available but machines can be expensive
X-Y plotter	out	good	slow	yes	paper free	yes	slow	fair to high

conclusion is that no one type is 'better', just more suitable for the current job. It should be noted that selection of the storage for a particular function often depends not on the most suitable type but on what is available at the time.

3.3.4 Input and output

The submission of data to and output from a computer, input and output or I/O, is achieved via a vast range of peripherals. The most common of these are summarised in Table 3.3 and graphics devices are examined in some detail in the next section. Once again, one tends to use the facilities available at the time, but a knowledge of the attributes of other devices often suggests methods of obtaining the best performance from existing peripherals.

3.4 Graphics

3.4.1 Data input

Graphical input to a computer must be translated from spatial relationships (recorded on maps, diagrams, photographs, held in tables of measurements, conveyed as digital information from recorders or mathematical descriptions of objects) into binary or numeric form. This is facilitated by a range of devices, the functions of which can be reduced to the following:

VECTOR DIGITISING. The recording of the relative or absolute coordinates of points on a chart or plan. This gives a vector record of the shape. Many devices are available to do this and details may be found in Newman and Sproull (1981).

SCANNING. Such as a television image or a remotely sensed data tape, where the whole image area is recorded by a string of values representing strips of single point PIXELS captured at regular intervals across the picture. This RASTER digitising technique is fairly new and arises from the advent of large amounts of cheap, fast data storage. An obvious example of a raster recording instrument is the television camera. Algorithms for translating between this and vector format are being studied because each format has different strengths and uses.

One more type of spatial data capture which should be mentioned is the recording of many parameters for a number of points fixed in space. The capture of this information can be accomplished by a logger, such as a weather station recorder, which produces input in a machine readable form or via a terminal, as typed input, direct to a suitable file that preserves the location of the source of the data.

No specialised graphical input facilities were present at Aston at the time of the project. The card reader or terminal were used for vector and raster capture as well as point data, with all the concomitant inaccuracy, inconvenience and waste of time associated with inappropriate data capture techniques.

3.4.2 Some input devices for interactive use

Interaction with a graphical image is accomplished via the usual alphanumeric keyboard and a number of special INPUT TOOLS. These tools are varied in type and function, some examples are as follows:

JOYSTICK, TRACKERBALL. These are linked potentiometers, mounted at right angles, which give X and Y coordinate information. They are often used to control the position of a CURSOR on the screen. Also INCHING BUTTONS, these are useful for specifying small X and Y increments.

MOUSE. A very ergonomic device consisting of two, X and Y, potentiometers driven by wheels on the bottom of the mouse and a set of function switches on the top. It can be used on any convenient flat surface.

TABLET. A drawing surface which detects the position of a pen/stylus and whether it is in contact with the surface or not.

LIGHT PEN. A photocell in a tube which gives a response when it 'sees' a particular picture element on the screen. A secondary function switch is usually provided.

TOUCH SENSITIVE screen. A grid of ultrasonic or infra-red transmitters/sensors across the VDU screen which work in the same way as the tablet but respond to the user's fingers rather than a special pen.

POTENTIOMETERS, FUNCTION SWITCHES. Simple dials and buttons which can be assigned meaning by the mainframe interactive program or the local intelligence of the terminal.

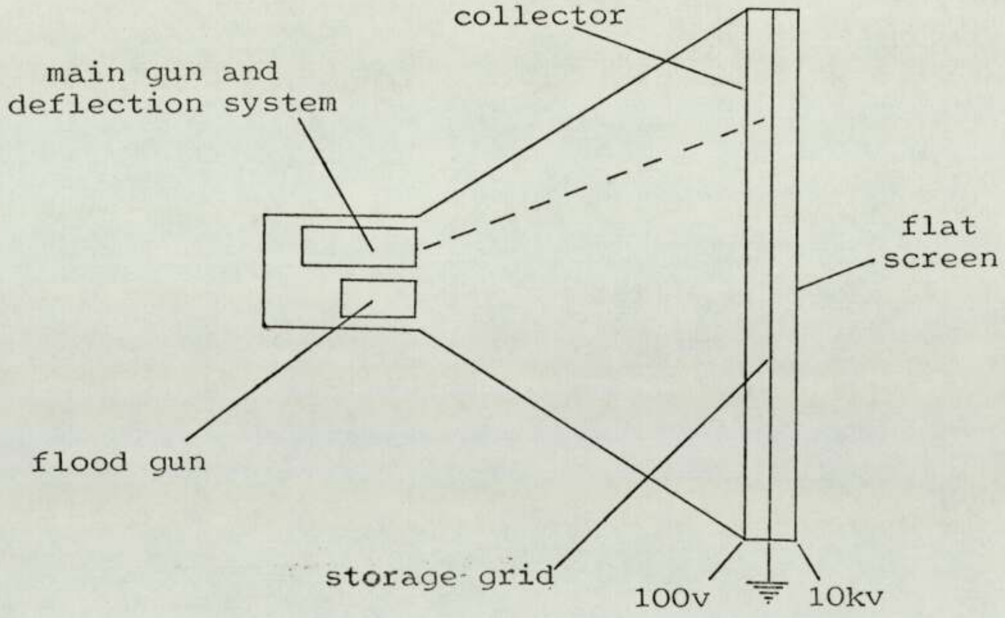
3.4.3 Temporary output devices

Graphical output peripherals may be divided into temporary and permanent display devices. Temporary images are almost always shown on a cathode ray tube (CRT). Three main techniques have developed and each type has different attributes.

3.4.3.1 direct view storage tube, Figure 3.1

The only manufacturers making this type of device are Tektronix Ltd.. Both these devices and the raster-based emulators are, for this reason, usually referred to as tektronix or TEK terminals. The image is 'drawn' on to a special long persistance screen phosphor and grid arrangement by the electron beam. The main problems with this device are that the image cannot be selectively erased and,because of electron leakage,the screen fades and a background glow builds up with time. This eventually renders the image invisible. The DVST uses standard ASCII codes (like an alphanumeric terminal) therefore it can easily double as an ordinary terminal. It displays a flicker free image, it is relatively cheap and the flat screen gives a very accurate image. The best way to consider a DVST in use is as a piece of paper with an indelible ink pen.

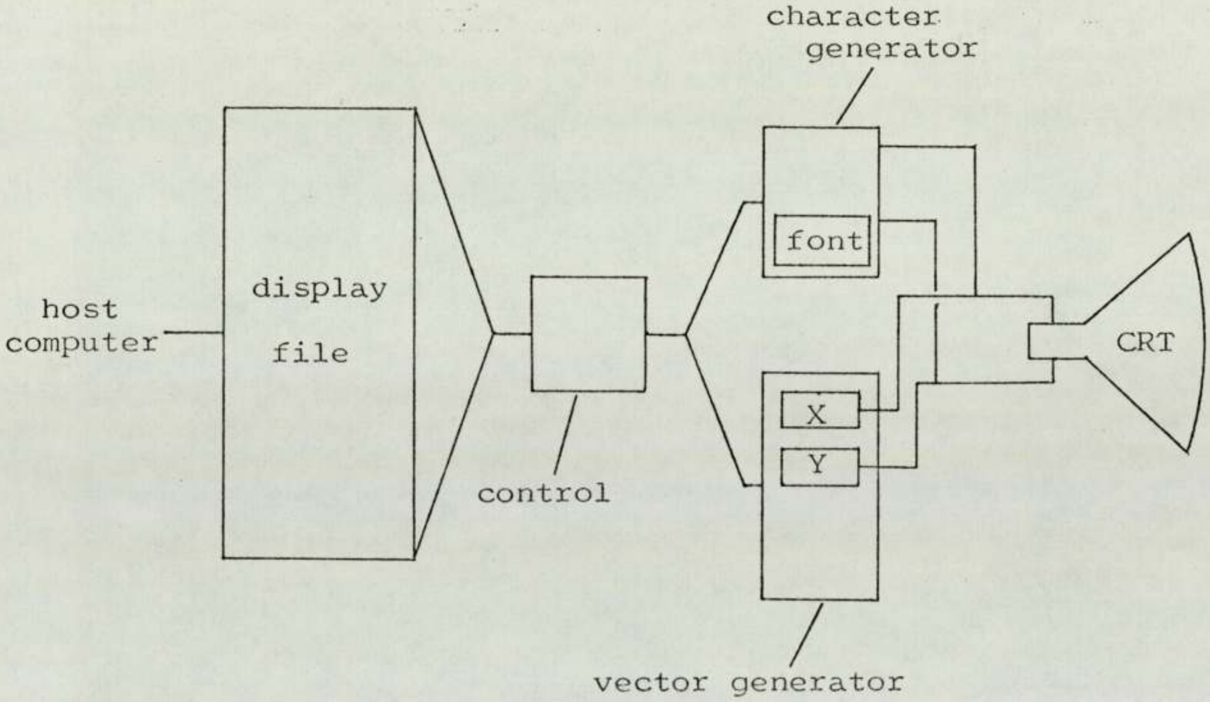
Figure 3.1 Direct View Storage Tube



3.4.3.2 refresh display, Figure 3.2

This device draws the image on to a normal CRT with the electron beam and continues to re-draw it before it fades away. The commands for the beam movement are held in a DISPLAY FILE. Obviously, there is a limit to the number of movements the beam, or PEN, can make in the 1/50th of a second between re-starting the image. The flicker caused by such image complexity overloading is the major problem of refresh displays. The advantage of this type of device is the ability to alter the image by simply changing the small display file. However, a special set of commands are required for such operations. The display file manipulation system often allows 'whole picture' transformations, such as rotation of a three dimensional image to simulate movement in real time. A useful analogy for a refresh display is a piece of paper with a pencil and eraser. This display allows the greatest and fastest interaction with the screen image. Colour is possible but is very expensive.

Figure 3.2 Simple Refresh Display

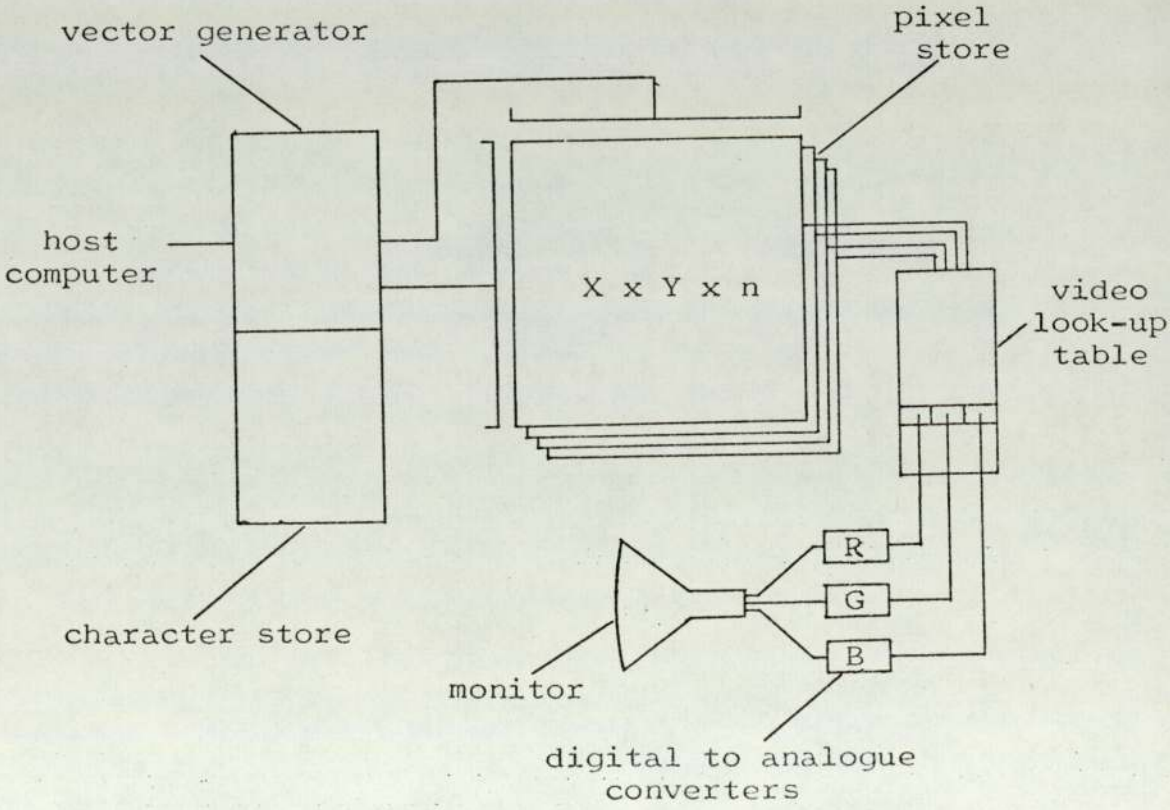


3.4.3.3 raster displays, Figure 3.3

These hold a matrix image of the screen in a fast digital store which is scanned in time with the raster on a standard television monitor. This BIT PLANE or PIXEL memory is read through a VIDEO LOOKUP TABLE (VLT) decoder which translates the binary number stored on the 'n' bit planes into an intensity value for the red, green and blue electron guns of the television tube. The bit plane can be written to, usually via a microprocessor control, by the computer. This type of display enables filled-in colour images of great realism to be created. Alteration of the picture entails overwriting of the previous information on the bit planes and is not especially fast. The definition of the image depends on the pixel, and thus the pixel store, size which, with the decreasing cost of digital storage, will soon be improved beyond the current industry standard matrix of 768 by 512 pixels.

Raster display is analogous to paper cutouts on a coloured background. The main problem with raster devices is the jagged appearance of the image which is made up of small blocks. There are methods of minimising this effect but they are neither cheap nor fast to implement and the facility for solid graphic images usually outweighs the disadvantage of this ALIASING effect.

Figure 3.3 Raster Display 4 bits per pixel and VLT



3.4.4 Hard copy devices

Devices which produce permanent output fall into two groups: vector and matrix. Vector based devices are typically single 'pen' plotters, either the accurate but slow FLAT BED or the fast DRUM. Drum plotters are generally provided as a service with mainframe computers. Flat bed types divide into small and cheap for user-accessible output on small (and micro) computers and large, very sophisticated, accurate map plotters. More exotic vector-driven output peripherals include CRT or laser recorders which produce fast output on photographic film and inkjet printers which give a very high quality of shaded copy, similar to airbrush work, on plain paper.

Matrix format devices are, like raster operation in general, becoming popular. This is because of the higher speeds possible with a large number of drawing elements in simultaneous operation and the implicit facility to create simple shaded images. The simple matrix printer is capable of producing very cheap stipple-shaded diagrams of low quality and, by using three coloured ribbons in succession, simple subtractive colour is possible. Electrostatic plotters offer higher resolution and greater speed than matrix printers, giving good shaded images. Finally, camera systems working on the TV raster principle can produce high quality coloured images on film or photographic paper.

3.4.5 Graphics libraries

All of the aforementioned peripherals have different characteristics when in use. The commands needed to carry out graphical or housekeeping operations vary with device standards, mode of connection to the host computer and the amount of local 'intelligence' of the device. Communication can be carried out in the local codes of the relevant machine. This is laborious for the programmer but minimises the amount of processing required and is therefore fast. A simpler method of interfacing the device to a program is to limit the graphical command sequences and calculations to a set of subroutines held in a library. These functions can be named mnemonically and included, supplied with the appropriate values for relevant parameters, in the user's programs when that particular operation is required. Most manufacturers provide such a LOW LEVEL library, written in a generally available language such as FORTRAN, for their devices. This library will allow all the functions of the graphics device to be used to the full. Examples of such libraries at Aston are the CalComp drum plotter routines, provided by the manufacturer, and the Sigma colour graphics library, written by Leicester University.

When a number of graphics units are required by one program or the program is to be used with different devices, possibly at different sites, there are considerable advantages in having a common set of functions which can be used with all the devices. Libraries have been produced that create a general purpose 'pseudo' code from the subroutines provided. They then translate this code, via a BACKEND routine, into device-specific code for the graphics unit currently connected. For communication with a different device all that is required is a call to a new backend. Because the graphics library is device-independent it can be developed to offer very sophisticated operations without fear of effort being wasted through device obsolescence. An example of such a HIGH LEVEL graphics library is the Computer Aided Design Centre's GINO-F package (1976) available at Aston. Around the kernel of GINO-F a set of sophisticated graphics processing routines have been created. Like GINO-F they are written in standard FORTRAN IV to allow the widest circulation. GINOGRAF facilitates graph plotting and graphical representations such as pie-charts and histograms. GINOSURF provides the ability to create and display three dimensional surfaces. GINOZONE is a digitiser processing package and choropleth map plotter executive.

3.5 Output Conflicts

Graphical output from a computer is a trade-off between the three factors of Speed, Cost and Definition.

The fastest devices are the graphics VDU, plus a camera for hard copy, and the electrostatic printer but these offer limited definition at any reasonable cost. The cheapest output routes, on a mainframe, are the alphanumeric VDU and line or matrix printers. Obviously, the definition of these devices is poor and the turnaround time of the printer is limited. The VDU has to display communications information as well as images,which causes further problems. High definition is obtained from film recorders, inkjet printers and good flat bed plotters but these devices are very expensive to buy and run. Flat bed plotters are usually used with photographic paper or film and the development time adds to the already considerable plotting time. Large DVSTs offer high accuracy of screen image but they are fairly expensive and have the aforementioned image persistence problems.

In general, allowing for the existence of a number of output routes, the device must be chosen to suit the requirements of the user and the data. There is no point in using a film recorder to produce perfect pictures of perspective viewed, hill shaded scenes if the surface is based on very few data points. Equally, if the map is required immediately a spline fitted high density topographic survey sent to the ECU for plotting accurate to one hundredth of a millimetre is not a great deal of use. If speed is paramount the user will often accept (qualified) lower quality, especially if the costs are low and in this situation the cheap, volatile CRT image should not be ignored. Quality always costs money and time but, if some effort is spent on definition-increasing photoreduction or some other, possibly manual, enhancements these overheads can be reduced to acceptable levels. No improvement can be made, however, if the original data do not justify the resolution required.

3.6 Considerations In Machine Selection

If a choice of computer facility is available to a user, or he is designing a new hardware system for a specific task, several factors should be born in mind in the selection of the best machine for the job.

3.6.1 Basic machine parameters

The tasks which the machine is required to perform can often be divided into batch or interactive. Most modern interactive computers support a background queue but old batch computers are rarely run interactively. If an old computer is used the obsolescence date and the expected useful life of the system created must be compared. The absolute size of the computer's core may also be important, especially in map data processing. However, most functions can be performed on small machines, provided that the unavoidable loss of speed can be tolerated.

Sufficient online and offline storage should be provided for all stages of the project. It should be noted that during development several copies of all the programs, all the input, test, intermediate and trial output data will require storage with fairly fast access time. Permanent erasure of a file of data will almost certainly guarantee that it is needed the next day due to a processing fault.

An inventory of the facilities provided should include the languages maintained, the packages and libraries and the peripherals available. The suitability of a language depends not only on its convenience for the project but also the speed, core requirement and enhancements of its implementation. If the necessary packages and libraries are not mounted the cost of purchase must be considered.

Transfer of information and programs in machine readable form may be required. The compatibility of the input devices with the available data should be ensured. The peripherals which are to be used during the project must be checked for their compatibility with the computer, specifically for the proposed operations. This depends on the speed, code set and software of both the device and the host computer.

3.6.2 Real speed

The effective speed of the computer to the user must be studied. A number of factors influence this real performance. The internal operations of the computer are controlled by a CLOCK. The clock of a mainframe computer will run up to ten times faster than that of a minicomputer. Microcomputers, in their turn, are between four and ten times slower than minicomputers. The efficiency of the mathematics is another influence on the overall speed. Once again, the modern mainframe is superior to the smaller mini or microcomputer. The average and peak number of users also influences the effective speed of a computer because each user takes a share of the whole system. This leads to a slowing of the 'turnaround', especially in processing and information retrieval.

The line speed of the peripherals which are to be used is often the critical factor in limiting the speed, especially in graphical display - because each picture may require a large amount of information to be transmitted to the terminal. For successful interaction the data transfer must be very fast, so that the 'conversation' takes place in real time, otherwise confusion occurs as to what response follows which action. If large amounts of data are manipulated the communication speed of the file management system in the computer should be considered. Finally, the amount of time that the computer is physically available to the user may also be relevant. This is especially important if many programs have to be developed in a short time or if much interactive work is planned.

3.6.3 Multiple machines and the current project

When a number of computers are accessible there is a possibility of using the one which is most appropriate for each successive task. For instance: production of a printed map requires a minicomputer for the digitising, a mainframe for synthesis and interpolation and a microprocessor controlled plotter for output. Or: many integrated science establishments now use a centralised small computer to capture and collate data for transmission to a large mainframe where it will be stored.

To give the current project the necessary input and output devices, and a future, initial input work will be carried out on the ICL 1904s. This was installed at the start of the project, has a card reader for bulk data input and Varey's 1977 datafiles are kept on this machine. The final system will be mounted on the Harris 500 to give fast interaction and access to the full range of graphics facilities. Archive storage will remain with the ICL 1904s until 1984...

4 STRUCTURE AND PHILOSOPHY OF THE DATABASE

4.1 A Three Dimensional Map In Computer Files

4.1.1 Definition of the map

A map is a set or sets of data which have a principally spatial relationship. This relationship is represented in the display of the data on a map or diagram. As long as the relationship is maintained on the diagrams produced from the data, and the data can be accessed by spatial definitions, the actual stored format of the data is irrelevant. The principal features which need to be maintained from the physical map are: the ease and speed of access to the information, the flexibility and suitability of the displays provided and the simplicity of use of the whole system.

4.1.2 Speed of the map

To meet these requirements with a computer based 'map' the machine must respond to the user as quickly as possible. The system has to supply both ephemeral and permanent displays of information. An interactive architecture for the host computer is the most suitable for this requirement. The capabilities of graphics output devices are required for the displays. This will allow an iterative approach to the retrieval and display of information. The user works with screen images rather than shuffling cards - the usual result of a batch system. To store the large amounts of interlinked data in the 'map' and allow fast access to any element DIRECT ACCESS, or 'disc', files will be used rather than the easier to work with but inefficient to access SERIAL, 'sequential' or 'card', files.

4.1.3 Appearance of the map

The appearance of the 'map' depends on the requirements of the viewer, or USER. For the user there is the need to see anything from single borehole records to whole maps of mixed parameters. All these must be presented in fairly recognisable forms which require the minimum of user education to be fully appreciated. The computer needs standardised, compact and structured sets of data which can be rapidly translated into any format that is required by the user.

4.1.4 Types of access to the map

The four types of access and retrieval that will be required are as follows:

4.1.4.1 maps

These contain a matrix of identical data elements for either all or at least a part of the mapped area. As such they require a large amount of information to be retrieved and displayed on a regular grid. The physical map can be either ISOLINE or ISOPLETH in form or may have the third dimension implied by, for instance, an ISOMETRIC plot.

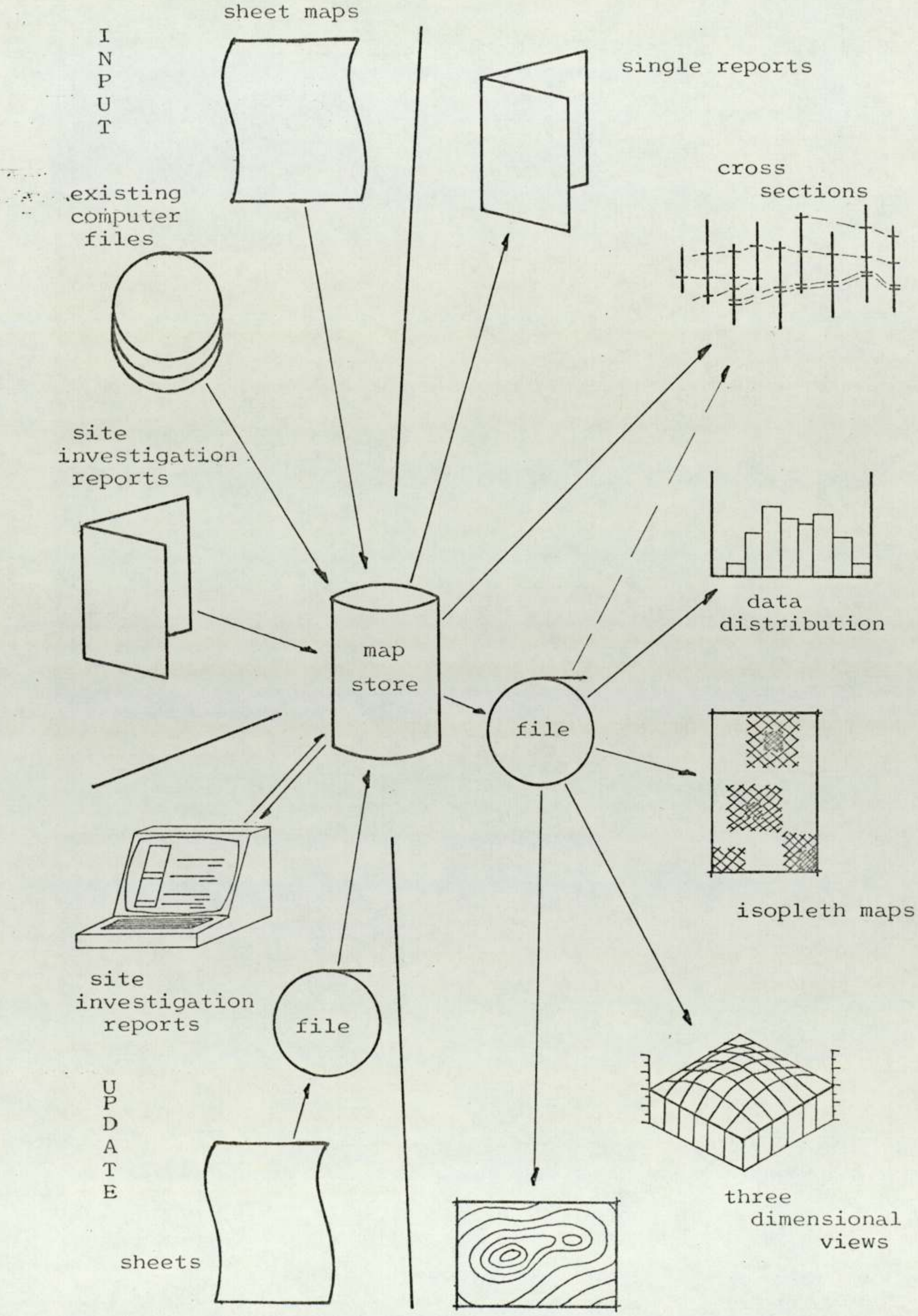
4.1.4.2 sections

Cross-sections of the map area, in any orientation or length, have to display limited information for a number of boreholes in any order of retrieval. This is probably the most complex retrieval and display problem faced. The facility to show both the continuous, e.g. levels, and single borehole data within a reasonable time and with enough detail to make geological inferences negates a simple structure for the storage files used.

4.1.4.3 segments

Whole, real or synthesised boreholes or interpolated

Figure 4.1 Simple Map Structure



records in an easily comprehensible form are the opposite of maps, as far as storage and retrieval are concerned. It is of advantage to be able to insert and edit at the segment level. This requires further complexity in the programs for this format.

4.1.4.4 hybrids

Beyond the types of display which are 'normal' for the geologist there will arise more sophisticated requirements for the data structure used,as the subtlety of which the displays are capable is realised. The arrangement of the files must be both simple and flexible enough to easily accomodate these extensions.

Figure 4.1 shows the overall structure of the 'map', to illustrate the above points and to summarise the requirements laid out in the second chapter. It also, incidentally, shows the general functions of any database.

4.2 Structure And Functions

4.2.1 Database concepts

Martin (1980) defines a database as 'a dynamically updatable body of data (a datastore) with its associated programs for input, output and housekeeping'. He then lists the eight fundamental operations which the programs must be able to perform as simply as if the user was handling a deck of cards, these are:

data manipulation

initial creation (ZERO) The making up of the original files.

addition / subtraction (+-) Insertion or removal of a record from the database, though not the files.

modification (MOD) Alteration of existing records.

selection (SEL) Retrieval and output of a record or set of records from the database.

restructuring (RES) After a number of +- operations the files may become cluttered with redundant records. This operation 'cleans up' the store.

data evolution

linkage of new files (+→) Attachment of more data storage files or new data types in different structures.

index creation (+I) Construction of 'maps' or indices of pointers to improve the efficiency of database operations.

addition of fields (+F) Placing new data in existing records of the datastore.

4.2.2 Possible storage file structures

The datastructure for this project has two conflicting requirements, the need to retrieve whole maps of one parameter and the recall of all the data on a particular segment. The single parameter or 'map' retrieval is a serial priority sequential reading process through a set of data. The best way to handle this requirement is to have a large number of direct access array files each containing one parameter or 'map'.

This suggests a very complicated indexing system, with sub files for each borehole 'segment'. It also implies great waste of space in some of the files. Giving the priority to the single segment reading will slow the recall of maps which could have more than one value per segment. But, in the 'map' structure the recall of all of the data for one segment would be a fearsome, and very expensive, task. Fewer files, each containing several parameters, will be required for segment priority storage and these will be easy to create, alter or update. The extension of the map or data types is also easy. There would be virtually no waste of space through redundancy.

Serial priority structure ('map' structure) has the advantages of being able to handle complex questions, because of the accessibility of every data element in the matrixed structure, and the fast retrieval of maps of parameter combinations. Against this it is difficult to set up or update. +- operations are not easy, frequent RES operations being necessary to keep the operation efficient. The size of the files and core buffers, with an associated machine overhead, would also be exorbitant for anything but the most rudimentary parameter set. This would lead to limitations on either the resolution possible or the size of map maintainable in online store.

From the requirements of the users of the final 'map' there is little need for fast access to many large area maps but a fair demand for fast access to site investigation type data such as single boreholes, sections and small maps. This requirement for speed with a lot of parameters, small file size and a simple structure suggests that the borehole or 'segment' priority structure would be the best to adopt, with a background, batch retrieval system and a separate, temporary fast storage method for 'map' style data arrays. The requirement for large area maps can be met by either pre-printed maps or the products of several pre-defined sets of information.

4.2.3 The chosen physical layout

To facilitate fast access to specific elements of data, save space wasted by the common absence of certain sets of values and simplify the decoding of records it is convenient to have generic data sets. Each fundamentally different set of data will be stored in a separate FILE, every RECORD in that file having the same FORMAT. These DATASETS can be linked by a priority structure of POINTERS which indicate the location of information in subordinate files. The use of these pointers by an enquiry program can cascade out all the required information for a particular segment, reading only the records appertaining to that segment.

The whole structure is given its spatial relationship by a master 'map index' array file which has the same proportions as the mapped area. The National Grid coordinates required are converted to index array addresses. These values are used to look up the pointer for the specified segment in a program array loaded with the current index file.

4.2.4 Operation of the database

The eight database functions can be performed on this datastructure as follows:

ZERO A program to read in the collected serial data, code and write to direct access files. The pointers and index are generated simultaneously.

+- Insertion of records at the ends of the relevant files and change the values of the pointers in the superior files.

MOD Rewrite the records in the relevant file, or replace all the segment's records in all the files as a set. There are no pointer value changes.

SEL Selection and reading of a record or string of records. Multiple choice selection of a whole map of data one parameter at a time. The mixing of these parameters to give new data analyses is carried out using secondary files held in a separate work area. Segment retrieval is fast, map extraction is relatively slow, because of the many criteria involved in the selection of the data. However, the processing of the maps in their relatively small 'map' files will be fast.

RES Copy all of the datafiles to new files with the pointers incrementing freely. The order of reading depends on the scanning of the map index. The final operation is to rewrite the index with the new pointers to the master or 'header' file.

+→ This operation would be performed if another set of data became available. For instance, sets of geophysical results or neutron activation logs. The replacement of the engineering data by a more detailed division of the test results was actually performed and is of this type. The header and new file are rewritten with the pointers for the new file placed in an empty or redundant field in the header file.

+I Index creation is not usually necessary. It entails looping a pointer through the header file, reading the check National Grid coordinates in this file,* calculating the array addresses of these values and writing the current pointer value to an array. Finally, the array is written out as the new map index. It is possible to create other indices for different priorities for the same files.

* All files must contain the National Grid coordinate, to facilitate repointering if a higher file is corrupted. The coordinate can be used as a simple check for corruption during normal operation.

+F If a new parameter or map of data is required to be added to the datastore the relevant files can be rewritten to accomodate this value in a redundant or newly created space.

Figure 4.2 shows the programs which implement these block operations and their interfaces with the datastore and output devices which make up the full database. The operations are divided into fast events that can be carried out in an interactive environment and slower tasks, more suited to batch processing. Another division is made between the single segment and the map operations.

4.3 Standardisation Of Formats

4.3.1 Structure of enquiry programs

The desirable features of storage and retrieval programs are that they should be:

LOOPABLE Decoding and insertion of information are made data-independent for compactness of the program. This allows the same read/write section to handle many different types of data on the same format.

FAIL-SAFE Anomalous values, such as no data or values out of range for particular parameters, should not cause catastrophic errors, such as a crash or datastore corruption.

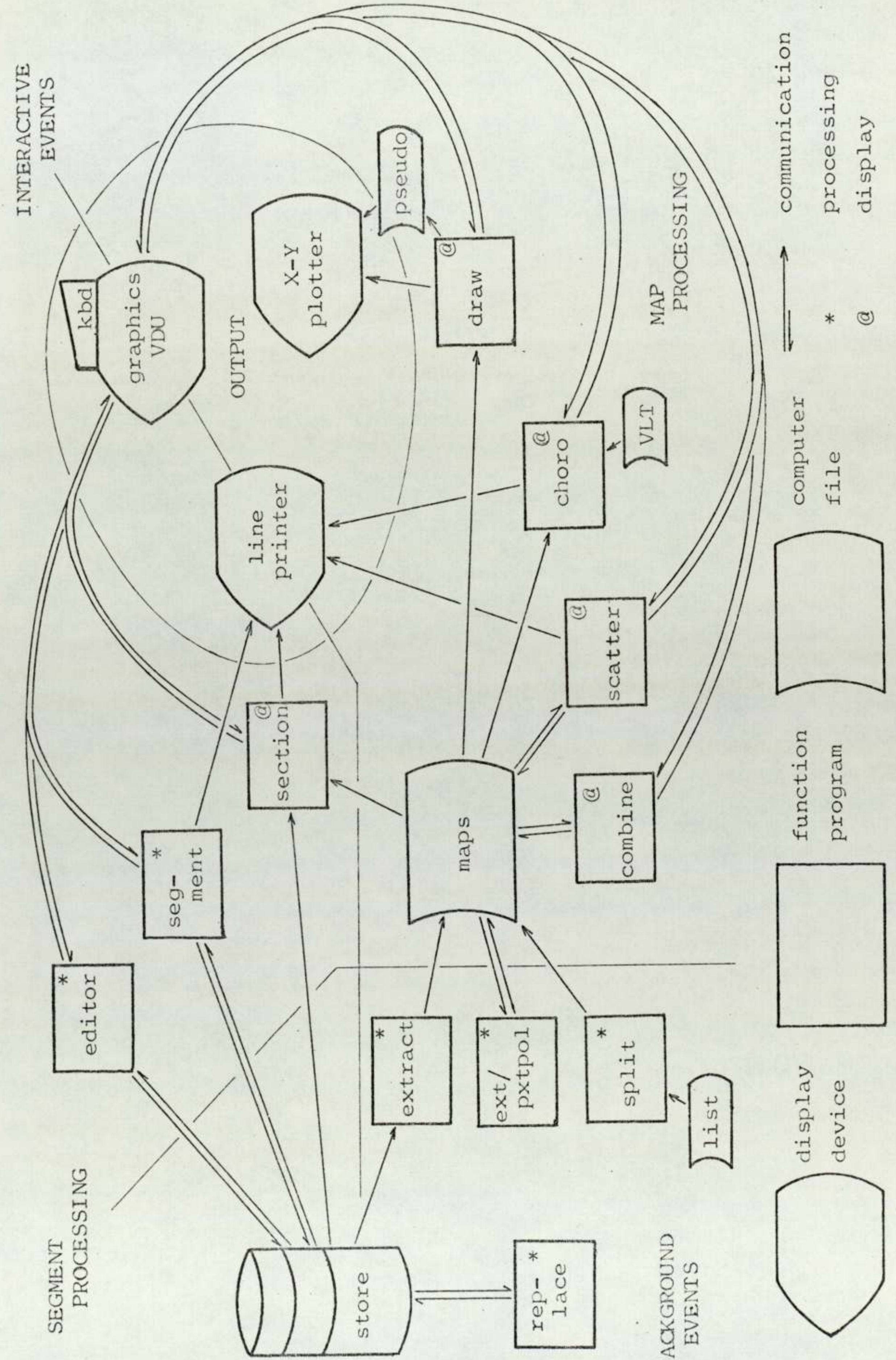
MODULAR Standardised input and output function blocks, uniform processing techniques in program subroutines and identical internal data formats facilitate the easy creation of display and manipulation programs.

4.3.2 Coding of the individual types of data

4.3.2.1 geological data

This is derived from the borehole log which is made up from observations on disturbed and undisturbed samples.

Figure 4.2 Full Schematic diagram of the System



It is usually written in a standard format, following the British Standards Institute (1957) or the Quarterly Journal of Engineering Geology Working Party Report (1970)* recommendations. Storage of this information on a computer may be achieved in three ways.

Direct storage of the original information as character strings. The advantage of this method is the easy storage and retrieval of the information. The disadvantages are the large amount of storage required for the extended text and the difficulty of processing the data by computer-aided techniques.

The storage of numeric or alphanumeric codes directly derived from the text descriptions, as in the IGS (Gover _et al_, 1971) or Cambridge LSD02 (Odell, 1977) systems, is much less expensive in storage and has the added advantages of simple use in the field, easy addition of new code types (under a generic classification scheme) and simple handling and selection of data types by interrogation programs.

Definition of types by the reduction to absolute parameter measurements. Examples are Varey's treatment of lithological types (by the specification of the percentages of the major constituent grain sizes) or the Hue Saturation Intensity colour classification system. This method is uniform in storage requirement and is thus economical for complex descriptions but not for simple types. Addition or recognition of new code types does not occur as an operation because the defining proportions are stored directly. Absolute parameter storage, or proportional storage, is not very compatible with species selection procedures, requiring a number of calculations to reach a 'type' for a stored

---

* Referred to from hereonin as WPR'70

value. But, it has a distinct advantage in display programs. Because of the human tendancy to name things rather than define them mathematically the stored representations (even if they have been captured before verbal classification) have to be translated into a readable form for text output. This destroys the advantages of this technique if it is used with anything other than graphical output. Coding of this format from the existing verbal description standards is clearly difficult for either the human or a computer program - the verbal and parametric systems being totally incompatible.

For this project's primary data sources the numeric code technique would appear to be the most suitable, allowing fast, efficient data capture and storage with simple retrieval decoding.

4.3.2.2 geotechnical data

These are given in the form of experimental results for occasional samples collected during the sinking of a borehole. In general very few tests are carried out regularly, only the Standard Penetration, moisture content, bulk density, shear strength and Atterberg limits being common enough to be statistically significant, Varey (1977). Except for the Standard Penetration Test (SPT) these are all given as real values. It is common for several readings for each parameter to be taken for one sample and several samples to be taken for one geological/lithological layer. The reliability of test data is not good, Bennett (1979). An averaging of the many results into a single value for each parameter from all the samples for each lithological layer is seen as a way of geologically locating the data and, at the same time, reducing the effects of collection and testing on the results. To back up this average parameter value the number of the associated layer, the number of measurements in the average and the most extreme value (indicating the spread of results or the variability of the sample/soil) are all stored.

Further consideration of the distribution and quality of the engineering data as well as the requirements of a 'worst case' map leads to the reduction of these records to a stratigraphic, rather than lithological, level of storage. Each set of engineering parameters are averaged for each lithological layer. Then, the worst value is selected from these to give a 'worst average' result for each age-related horizon. It should be noted, however, that the datastructure employed would be equally compatible with whole samples, lithological or, as initially considered, whole-borehole worst case values.

4.3.2.3 archive, chemical and level data

Several parameters have, or can be averaged to, only one value per segment. Because these are fairly few in number and are required frequently it is convenient to store them in the master file containing the pointer information for the subordinate files (which hold several sets of data for each segment). If many parameters were mapped to the spatial index in this way an electively-read secondary file would be used to reduce the number of possibly redundant values read along with the pointers and thus save read/write time.

The types of information held in this file include: the worst sulphate concentration and pH for the boreholes in the segment, codes indicating the presence of man made hazards, the average ground level and the worst overall SPT.

4.3.3 Indication of reliability

The advantages of averaging all the borehole records on to a grid of segments and reducing the engineering properties have already been mentioned: reduction of errors, reduction of storage size, ease of handling, statistics of interpolation. As a result of this averaging there is a danger that the final synthesised boreholes (or SEGMENTS), whether real or interpolated, will appear homogeneously accurate and reliable. To warn the user of the level of trust to place in the presented information it is important to maintain a register of either the number of real measurements which make up each parameter's value or the separation from real data for interpolated parameters, as an index of reliability. The more extreme measurements for average values should also be given, as 'error bars' on those data.

4.3.4 Data processing

The arrangement of the values for individual parameters inside the files of the datastore is influenced by two factors. The size available to the store and the speed of access required. In fact, the store requirement should always be minimised regardless of the space that can be used for economy, ease of transfer, space for expansion etc..

Database operations require a large volume of data transfer between the central processor of the computer and the disc storage. This is usually the limiting factor on the speed of operation of a system. To reduce this overhead both the number of records read and their length should be minimised. Because the whole record is copied into an operating system buffer (the rate determining step), even if only the first word is read by the program, care should be taken to program for a minimum of transfers.

The length of records can be reduced by PACKING as much information as possible into each WORD of the record. This packing has an associated overhead in code and decode calculations, but generally this is small compared with the time saved in information transfer. However, it is worthwhile ensuring that the most commonly required data have the minimum of decode time.

The method use to pack several integer codes for different parameters in the same computer word for storage is simple. The values are multiplied by factors of a base, in this case ten, and added together, i.e.:

word = value 1 + (value 2)X10 + (value 3)X1000

where, value 1 is one character wide, value 2 is two wide and value 3 can be up to the maximum integer width on the machine minus the three already used. (e.g. Harris, 7characters minus three = five)

Real variables, such as the engineering data or levels O.D., can be converted to integers by extraction of the significant figures. This is achieved by a standard multiplication or division by factors of ten. These factors are recorded for re-insertion by the decoding programs. It should be noted that this method limits the resolution of the real variables to the field width allowed. But this can be tuned to the accuracy of the parameter.

The full encoding process for the making up of a record ready for storage is as follows:

integerise real values

check the width of real values to ensure that the data
fields on either side are not corrupted by stray

characters

multiply up and sum the absolute values for parameters for
each word in the record

carry out any transfers of mathematical sign necessary

buffer out the record as a string of variables or an array.

Decoding consists of reading the record into a buffer and using successive remaindering to calculate the values of the original parameters in each field. This process can be made more efficient by standardising field widths and using loop decodes to write the results to successive array elements.

Packing the records in the files leads to an effective format in these files which can be likened to the format of serial files. Unlike serial files, direct access files cannot be read directly with specified field types and sizes, or FORMAT STATEMENTS, because they are stored as binary codes, not characters. However, it is convenient to think of direct access files as having parameter fields. These formats and the numeric codes used for encoding the information captured are shown in Figure 4.3 and listed in Appendix 2. It should be noted that the coding employed permits generic searches and comparison of the data, as well as the definition of specific parameter values. This is one of the advantages of the numeric coding of data - these generic searches can even be made on packed data.

4.3.5 Efficiency of the packing

For ease of transfer, visual inspection, hard copy or machine readable backup it is useful to list the contents of a direct access file to a serial file. It is possible to fully decode the packed file during this operation and to format the output meaningfully, but this may cover up some

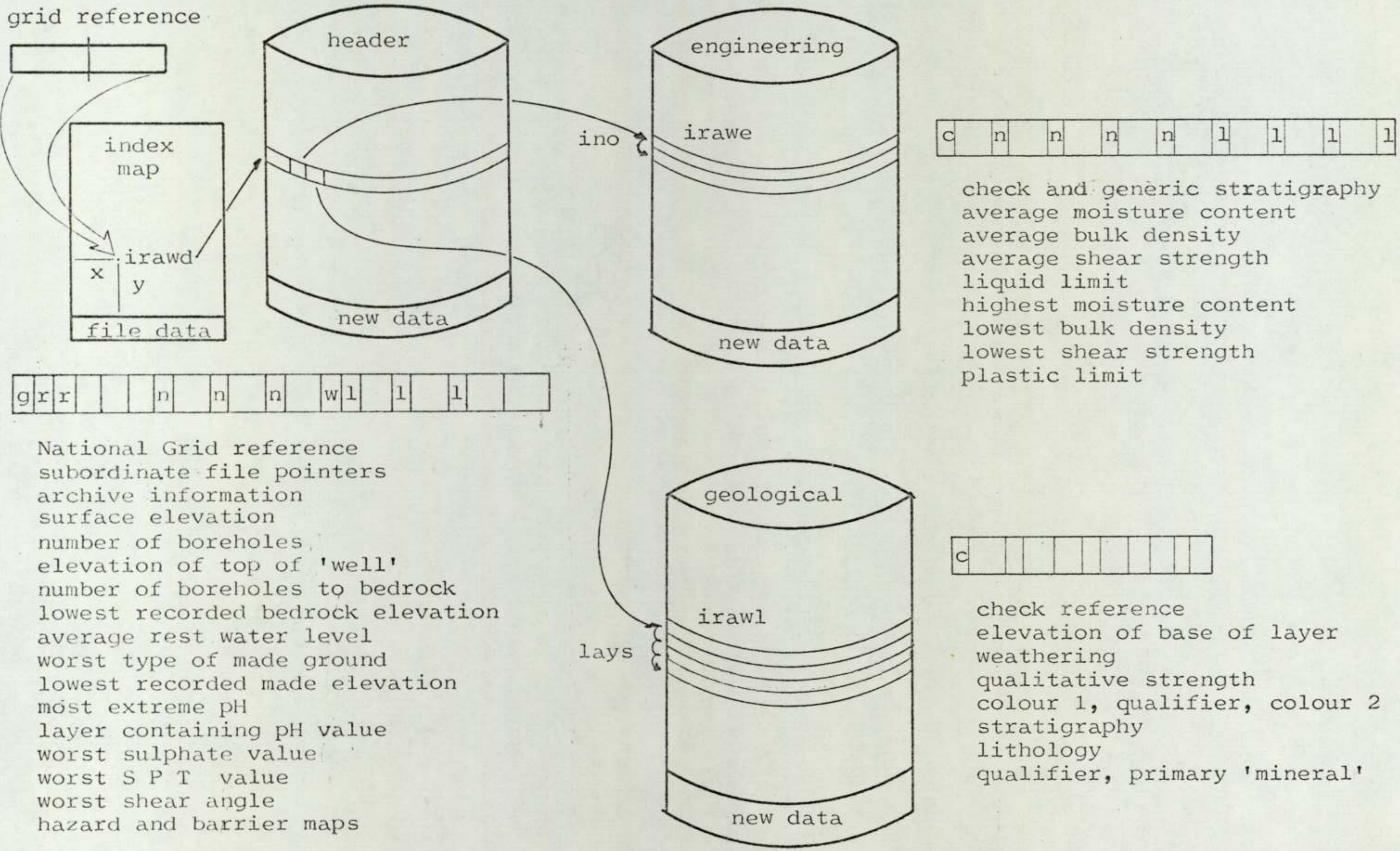


Figure 4.3 Filestore Structure

interference problems between parameter fields and will certainly increase the size of the output file. This relative size of direct access and serial files depends upon the packing of the direct access file and the format of the serial file. With suitable packing the direct access file is capable of better numeric storage efficiency than the serial file. For instance, in this project the serial listings of the raw data and full datastores (Appendix 6 fiche) are half as large again as their direct access sources, even with a compact eight character wide integer format for the output files.

4.3.6 'Map' temporary storage files

The format of the temporary storage files for the extracted 'map' arrays of data should be compact enough to read and write quickly. This suggests that the values should be integerised to save space. The file, however, must allow the same accuracy of data as the main datastore. The only parameter measured to more than four significant figures is the level of the top of the boreholes. The averaging to 100 metre square segments obviously greatly reduces this resolution and therefore four significant figures was chosen as the minimum resolution of the system. A different problem occurs with the sulphate concentrations or intermediate calculation maps, where the range of values is greater than four significant figures. In the interests of speed and compactness this loss of accuracy has to be tolerated and, of course, logarithmic coding could always be used if range were more important than resolution.

The maximum number of characters in a serial file record on the Harris 500 is 132. Every row of an (integerised) four characters per segment will therefore fit on two lines of a file. The integerising of the variable to enable maximum resolution can be standard or user-controlled. The SCALE FACTOR is recorded at the start of the map file for use in the re-conversion of the data. Other file specifications are also useful - the area covered and an indication of the type of data (real, integer or processed). Trailing information, on the classes into which the data have been divided or decoding information, may also be added to the map.

4.4 Realistic Use Of The Database

4.4.1 Practical environment

So far the theory of a spatial database has been presented. This structure has now to be implemented on a real mainframe computer in a practical multiuser environment. The Harris 500 has been selected as the device to be used. Its limitations are as follows:

STORAGE The machine loading from other users considerably slows the transfer of information to and from the disc filestore. There is no magnetic tape archive provided for general users. The filestore allocation for users is limited. Therefore, care must be taken to make disc transfers efficient and to keep the datastore, programs and peripheral data files as compact as possible.

COMPUTING POWER The Harris is a very fast modern machine, with virtual memory. This effectively means that the core size can be extended by 'page swapping' to a maximum of two Mbits. The result is that, with care, there are no machine limits for the current project. The terminal speed is 9,600 baud. This is high for normal use but places some limits on the interactive capabilities of the graphics devices, which are not closely coupled but connected as terminals. Thus the same precautions of data transfer efficiency should be observed for graphics as for disc transfers.

RELIABILITY This is made up of two parts: program and machine. The machine is unreliable, having as much as 25% down time per week. This occurs in a series of short, unpredictable crashes rather than long periods. Fortunately, data corruption is rare. In the time available it is not possible to construct a totally foolproof system of programs. This leads to a dependence on the 'break' key of the terminal for crashing and restarting programs from the beginning if user errors occur. But, on the Harris, the system performance prevents either of these halts from being too time consuming. Re-running of the programs is an adequate solution for the present. Exceptions to the 'break' and run process are the creation, or 'zero', processes. These run for very long periods with precious data and have to be dumped as a protection against the loss of file pointers in the case of any failure.

4.4.2 Resultant programs

The programs for the database are of three types:

CREATION. The zero event. A set of programs to collate, correct and store the captured information. These are both batch and interactive and are based on the ICL and Harris machines. The common factors are complexity, reliability and one-shot operation. These programs are data and system specific and have not been included in the final system for, like most 'real time' processing, they are not generally usable. Recommendations are made for improvements in data capture techniques, including some techniques from the creation programs in later chapters and Appendix 3.

INTERACTIVE programs. Operations dealing with the selection of a few input, output, graphics display and "hands on" manipulation of maps. All require the user to converse with the operating program. These are run from VDUs or graphics devices. The user-access to the graphics terminals is limited and facilities for the majority of the programs to supply meaningful output without recourse to sophisticated presentation devices will increase the usability of the system.

BATCH or background programs. In an ideal world all the operations would be immediate and run interactively. Unfortunately the amount of processing required by some database operations is antisocial in a multiuser environment. These programs have to be run as slower jobs, which are carried out by the machine when time is available. The jobs are submitted to a queue, suitable parameters are supplied and the programs usually run within a day. Operations carried out in this way are complex map processing, involving intersegment operations, and the whole store processes: 'map' extraction and the RES function.

4.5 Program Algorithms

4.5.1 Methodology of construction

It has already been mentioned that the use of standard formats for data simplifies the processing operations. This can be extended to the programs for manipulating these data by employing standard read/write routines, display functions and program structures. These common function blocks are connected into a few flexible programs which can execute a wide range of options, as selected by electives provided at run time. This technique facilitates fast programming because only the new segments have to be debugged. It also reduces the amount of storage required because the few programs can be held in a compiled form and are instantly available to the user.

4.5.2 Requirements for the language used

The computer programming language used to write the system has to fulfill several requirements:

GRAPHICS There is a need for compatibility with the graphics libraries provided and a logical structure suitable for interactive operation.

LIBRARIES Many programs will have common sections, at least in interrogation and display. In addition, program construction is eased by the use of function routines. To support this, and the graphics libraries, the facility to simply handle many subroutines is necessary.

ARRAYS Map files and data for single segments are held on matrix structures for loop processing and ease of handling. This makes for compact programs and flexible manipulation. Array operations must be easy to use and fast in operation.

TEXT The soliciting of electives by conversational displays and the translation of the coded information into an understandable form requires good text handling facilities in the language used. Thus, string comparison and display with compact text storage and simple output formatting should be provided.

4.5.3 Choice of the language

This project was started in 1979 and the only language available at that time which was fully compatible with the graphics libraries was FORTRAN 66. This language also has good file and array handling facilities. FORTRAN 66 is widespread in the academic and commercial environments with a concomitant high level of awareness in the standards and use of the language. It was therefore selected as the common language for the system.

Nowadays the situation has changed markedly. With the increasing importance of microcomputers older, faster and more compact languages are being revived. The best of these, PASCAL, is being implemented on mainframes as well as micros because of the simplicity, elegance and economy of its logic. PASCAL is highly suitable for the structure of interactive and graphics programming. The text handling is good, alien libraries may be used and it functions well in a multiuser environment. If this system is re-implemented PASCAL is recommended as the language to be used.

4.5.4 Types of program

There are three main classes of program: retrieval, processing and display. One program may do all three as parts of its operation, here it will be classified according to the first stage carried out. Schematic algorithms are given to illustrate the broad logic of the programs.

4.5.4.1 retrieval

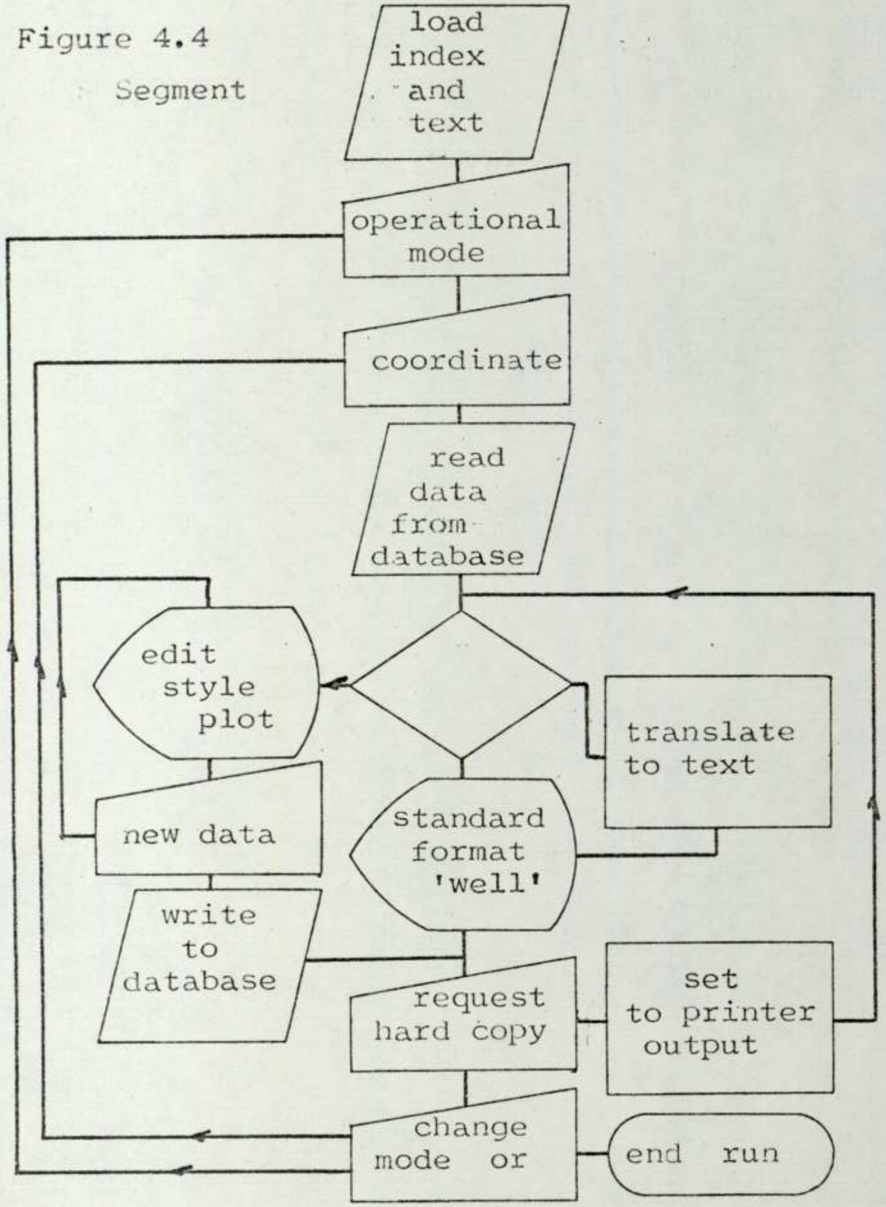
Three types of storage interface are allowed:

SEGMENT Figure 4.4 Displays and allows insertion of single segment information via a normal terminal with line printer hard copy.

also-EDITOR Figure 4.5 A housekeeping program. Displays a segment on the colour graphics terminal, following the recommendations of Bouma and Nota (1961), and has the

Figure 4.4

Segment



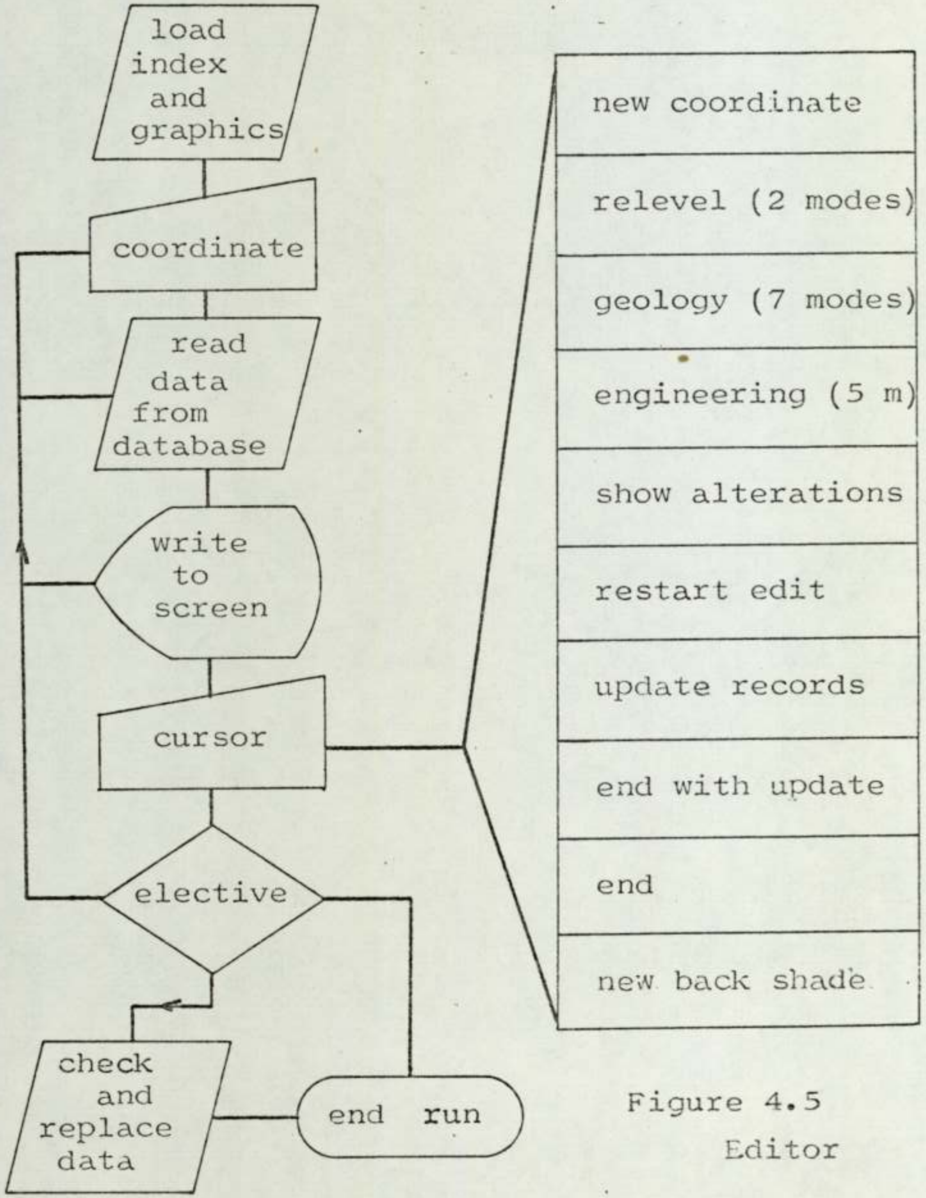


Figure 4.5

Editor

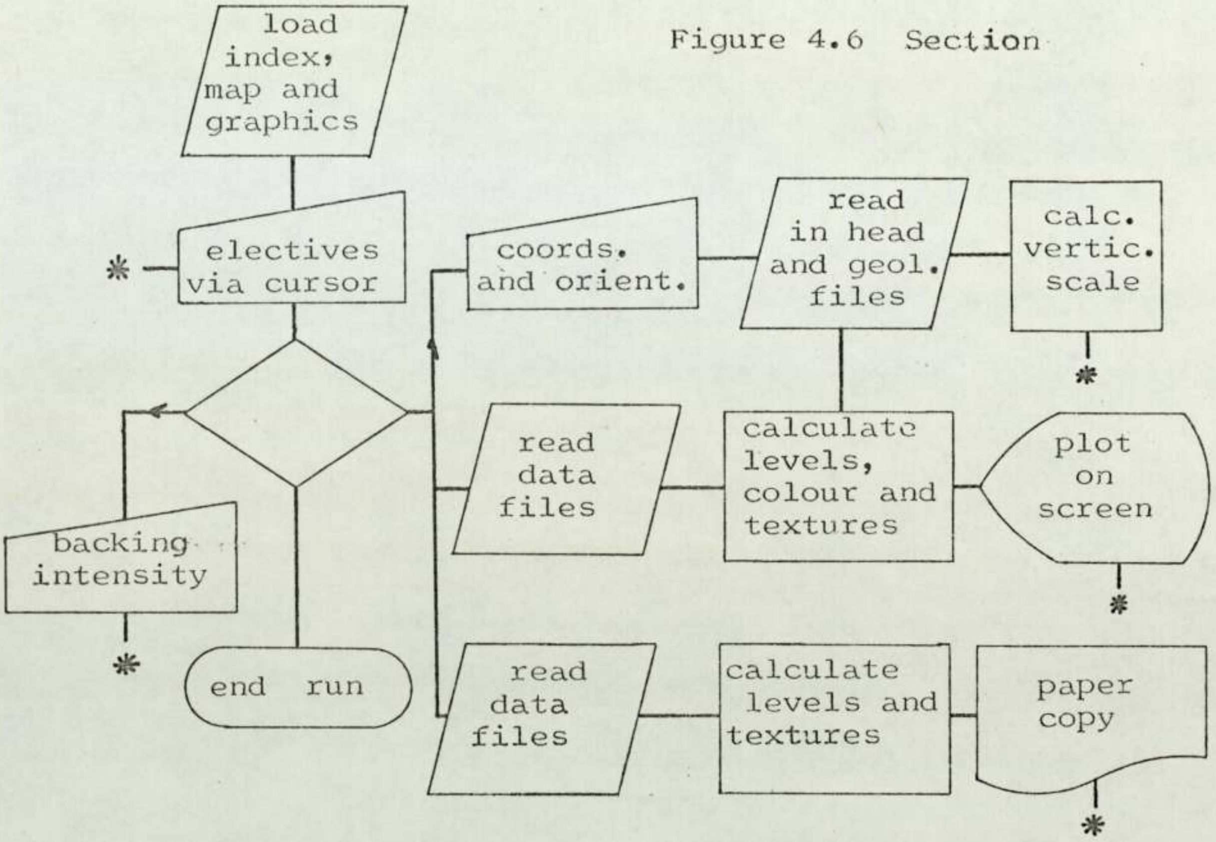


Figure 4.6 Section

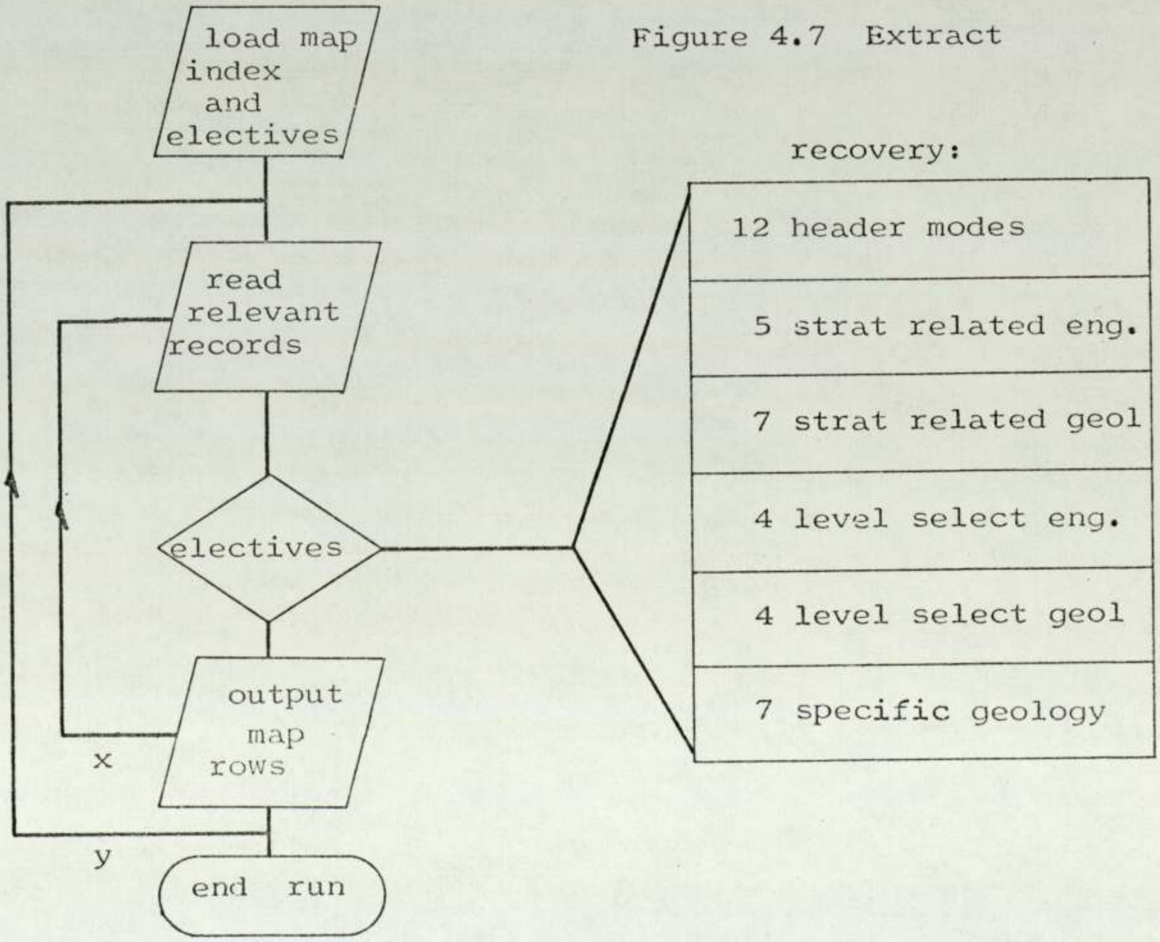


Figure 4.7 Extract

facility to conversationally alter elements of the data.

SECTION Figure 4.6 Draws sections across the mapped area in full colour and with texture, with the facility to make line printer copies of these sections.

EXTRACT Figure 4.7 Allows selection of maps of data from the datastore.

4.5.4.2 processing

Five programs for the manipulation of the map arrays of data are provided:

COMBINE Figure 4.8 Facilitates the integration or separation of maps. It is a general purpose read/store/write loop with many elective tools.

SCATTER Figure 4.9 Allows the division of map data into meaningful classes for display.

EXTPOL and PXTPOL Figure 4.10 Interpolation routines for real and integer data respectively. They provide a simultaneous reliability map.

REPLACE Figure 4.11 This is the one updating program for the RES function, it can also be used to store new maps in the datastore.

4.5.4.3 map display

DRAW Figure 4.12 Offers isoline, isopleth and isometric plots to give a high quality line drawing capability on graphics VDU or plotters via the GINO-F library.

CHORO Figure 4.13 Provided to make use of the choropleth mapping capabilities of the colour graphics display and to produce fast line printer maps from the system.

4.5.5 Statistics of the programs

Some data for these routines, in FORTRAN on the Harris (32 bit words, 1 sector 2688 bits), are shown in Table 4.1.

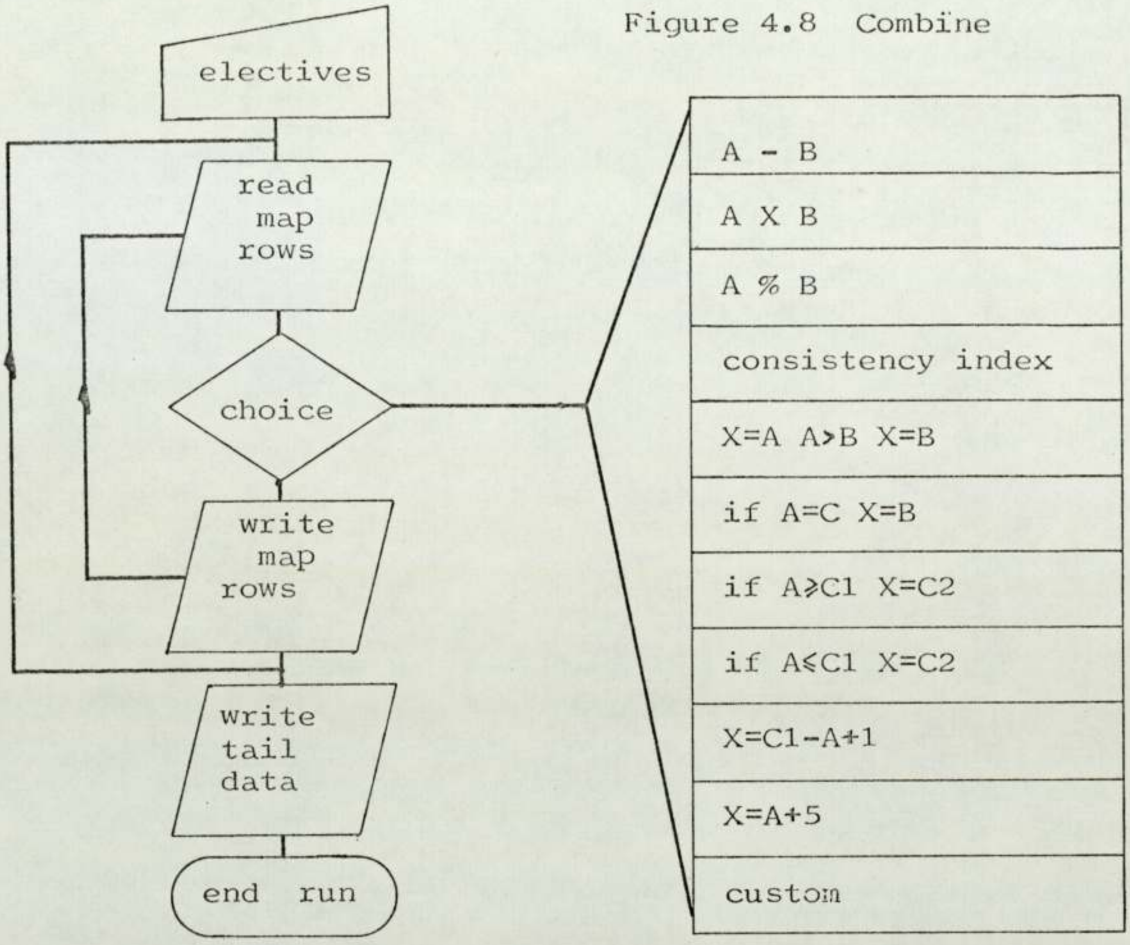


Figure 4.8 Combine

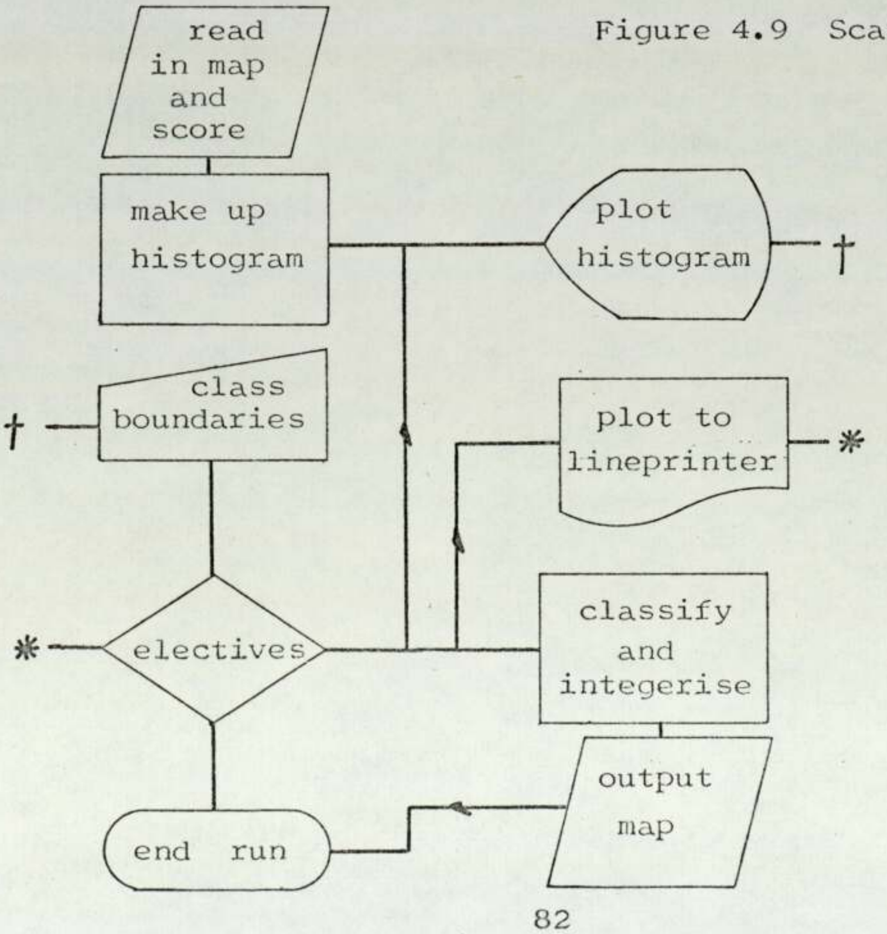


Figure 4.9 Scatter

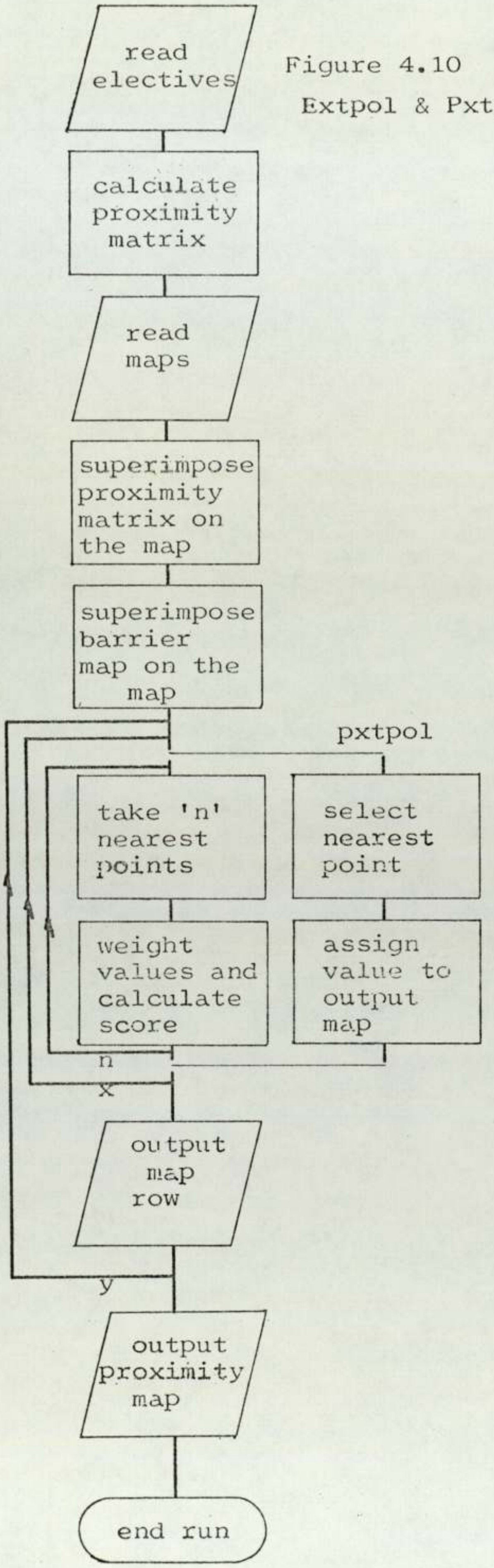


Figure 4.10 Extpol & Pxtpol

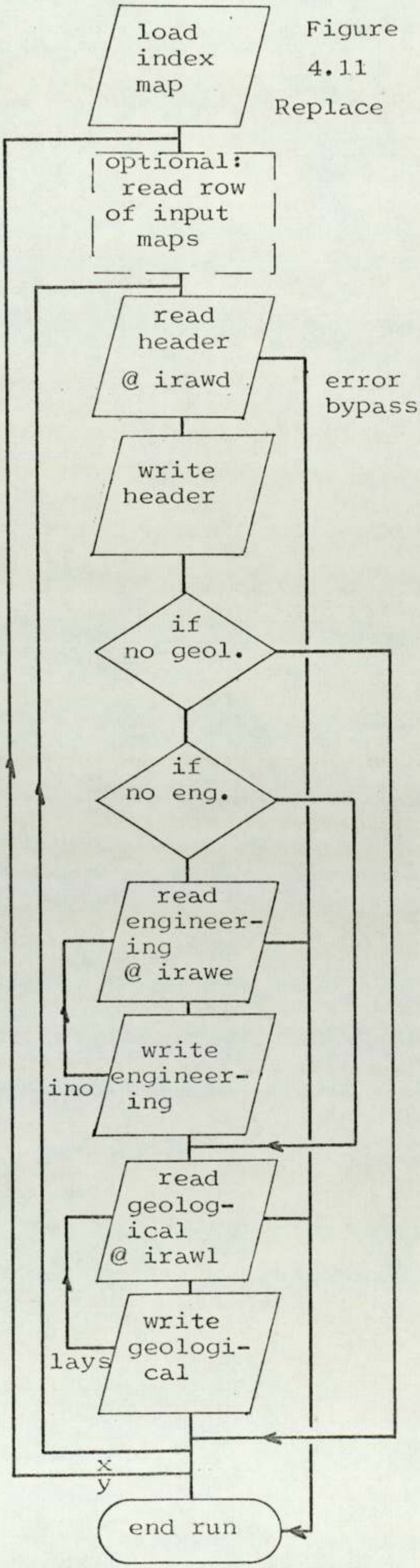


Figure 4.11 Replace

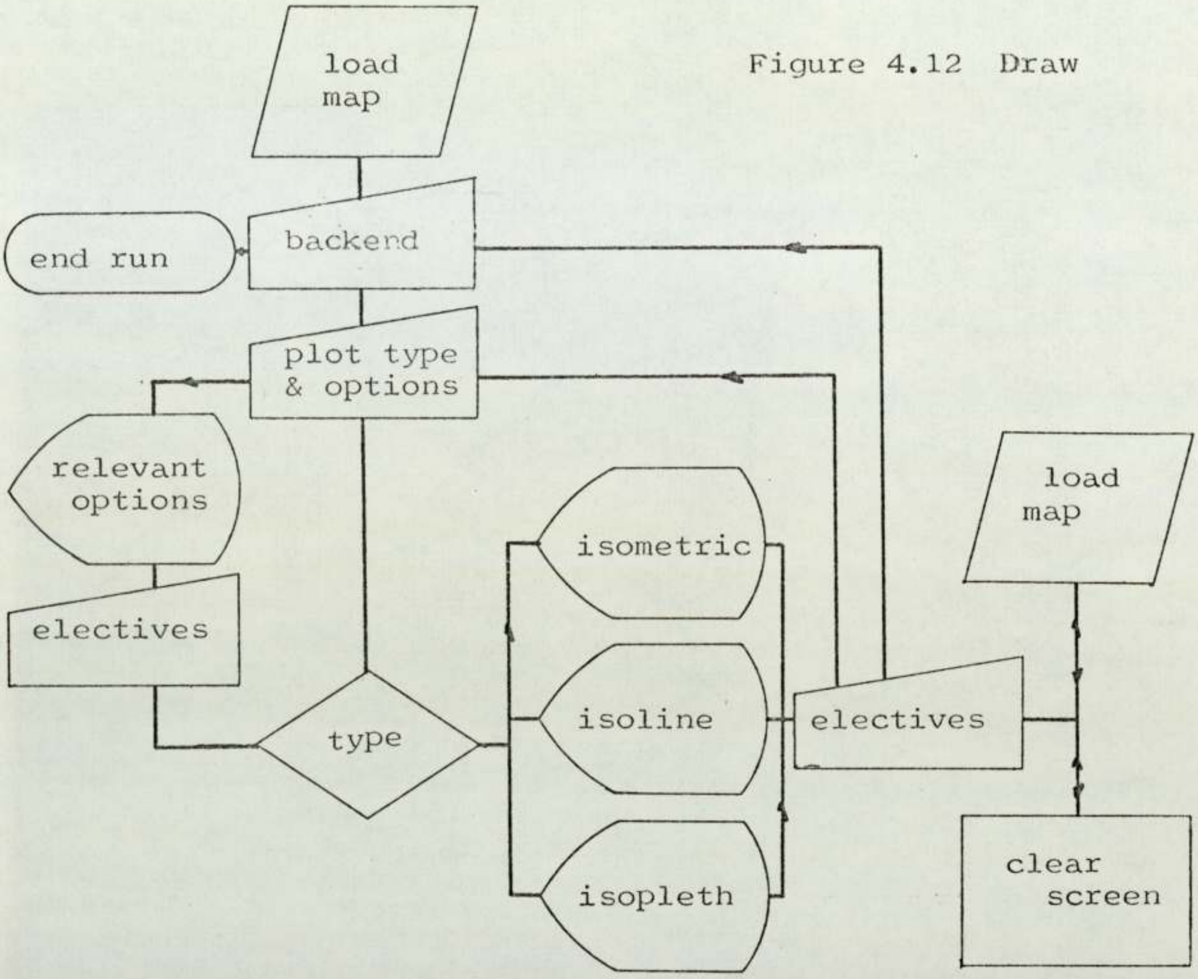


Figure 4.12 Draw

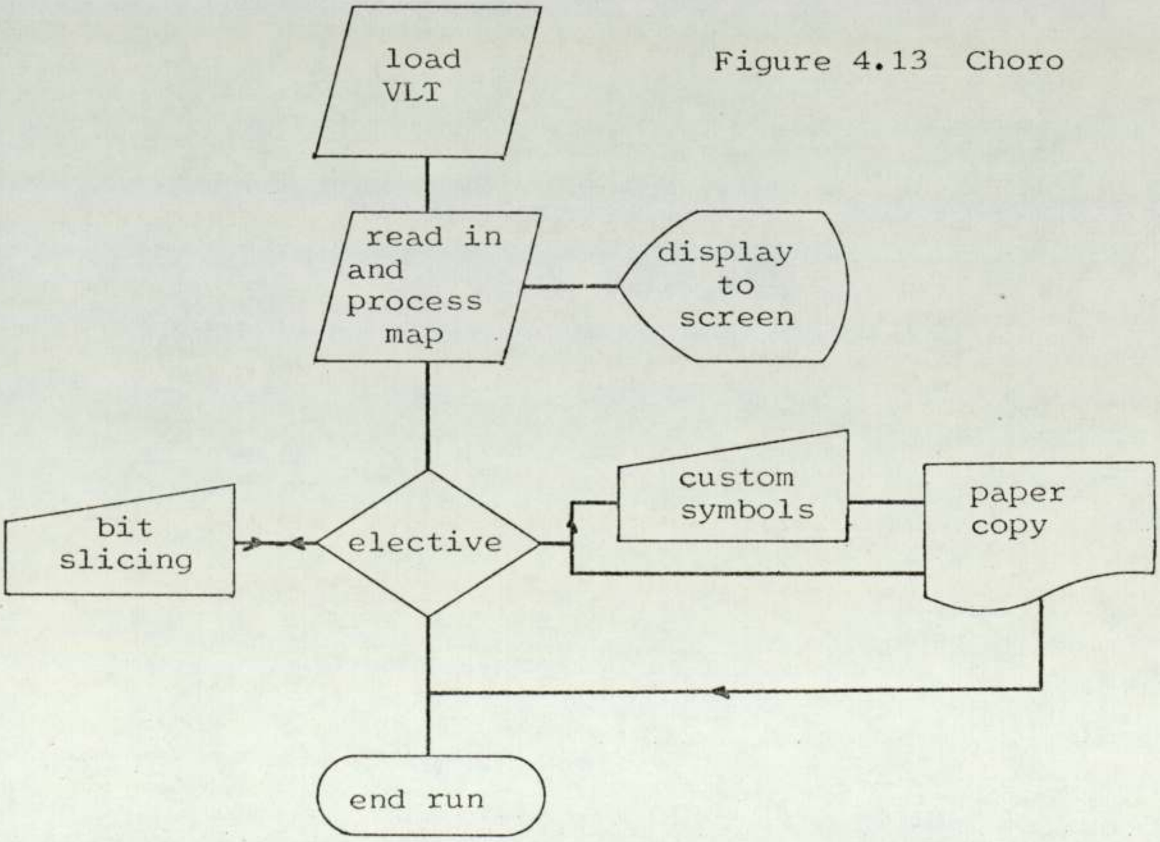


Figure 4.13 Choro

Table 4.1 Statistics of the Database Programs

PROGRAM		compiled core KBits	uncompiled store Sectors	run time seconds CPU time	notes
EXTRACT		26½	48	50-80	time depends on number of out maps, dump 7 mins.
EXT/PXTPOL		57½	16	180-540	per map depends on the number of points given
REPLACE		22	16	180 or 420	streamline or restore
COMBINE		13½	24	10 + 5n	first and 'n' subsequent maps input
SCATTER		37½	16	6•5	includes a hard copy listing
CHORO		27½	16	8•7	includes a hard copy and some bit slicing
DRAW	isometric	245½	40	16	
	isoline			25	compile time is 21 sec. extra, once per session
	isopleth			50	(for a full screen of information)
SECTION		50½	56	28	for a 40 wide section plus rescale plus hard copy
SEGMENT		33	40	3•5	for start, one 'well' and hard copy
STARTUP		–	16	45	compiles all main programs once per session
PSOUT		–	8	240	for stored map output
EDITOR		41	72	1•5	per 'well' displayed, compile is 20 sec extra

5 GRAPHICAL TECHNIQUES AND STANDARDS

computer graphics in geology

5.1 Drawing By Computer

This project relies heavily on the use of the digital computer to produce displays of geological information in non-textual form. The hardware and software necessary for this function have already been described. This chapter is an explanation of the techniques in Computer Graphics which have been applied to the interactive display of the data collected.

5.1.1 Graphical output devices

The commonest way of producing a diagram or picture is to drive a drawing instrument, a PEN, over a surface, raising the pen to invisibly MOVE and lowering it to visibly draw LINES on that surface, to make up the desired IMAGE. The technique is the same manually or mechanically. It gives a high definition image which is relatively slow to produce and requires specialised VECTOR commands. In computer-aided drawing these commands are issued via a LIBRARY of routines which take care of much of the translation and complexity of the codes involved, leaving the programmer free to order moves, lines or drawn TEXT strings irrespective of the device used.

5.1.2 Graphics packages

More sophisticated routines are provided via HIGH LEVEL graphics packages such as GHOST or GINO-F which provide the ability to command: circles and arcs, windowing (trimming to fit a specified area of the picture), image transformations (scaling, shearing and rotations), drawing and viewing in three dimensions, the simple plotting of graphs and charts and the facility to divide a picture up into a number of segments that can be manipulated independently. When using vector graphics it is important to remember that the whole image is created by the movement of one pen. The more efficiently this pen is moved the faster and cheaper the image will be produced.

5.1.3 Non-graphics devices

Printing devices, such as line printers, also have some graphics capacity. On some devices the typeface or FONT can be changed to extend the implications of the characters. In addition the spatial relationships of the characters on the page can, despite the minimum resolution of the character box, be employed for images. The advantage of using printers for graphics is in the availability, cheapness and speed of the devices. The printer has one major disadvantage as a plotter, the movements of the paper and type head are not reversible. To overcome this problem the output to the device must be ordered from the top left to the bottom right of the image.

Visual display units (VDUs) have all the problems of the printer plus a reduced drawing area and no chance to improve the final definition by photoreduction of several pages of image. They are, however, faster than any other device and are universally available.

5.1.4 Temporary output devices

The graphics terminals, which were reviewed in an earlier chapter, offer similar speeds to the alphanumeric VDU but far higher definition. They are vector plotting for line drawing with a variety of text output methods.

Each type of terminal has characteristics suited to different applications in geology. The direct view storage tube (DVST) is appropriate for high quality diagrams such as contour maps or other stationary monochrome representations. Raster displays can give 'real' representations of colour and texture such as three dimensional solid structural diagrams or partially animated cartoons, for example of vulcanogenic processes. Refreshed picture devices allow man and computer interactions with the screen image in real time. This facilitates operations such as the display of the genesis of a species as an animated line drawing or could provide the ability to use a light pen to sketch sections through boreholes displayed on the screen.

5.2 Representation Of Three Dimensions

The alteration of a variable with its location in two dimensions can be displayed in a number of ways:

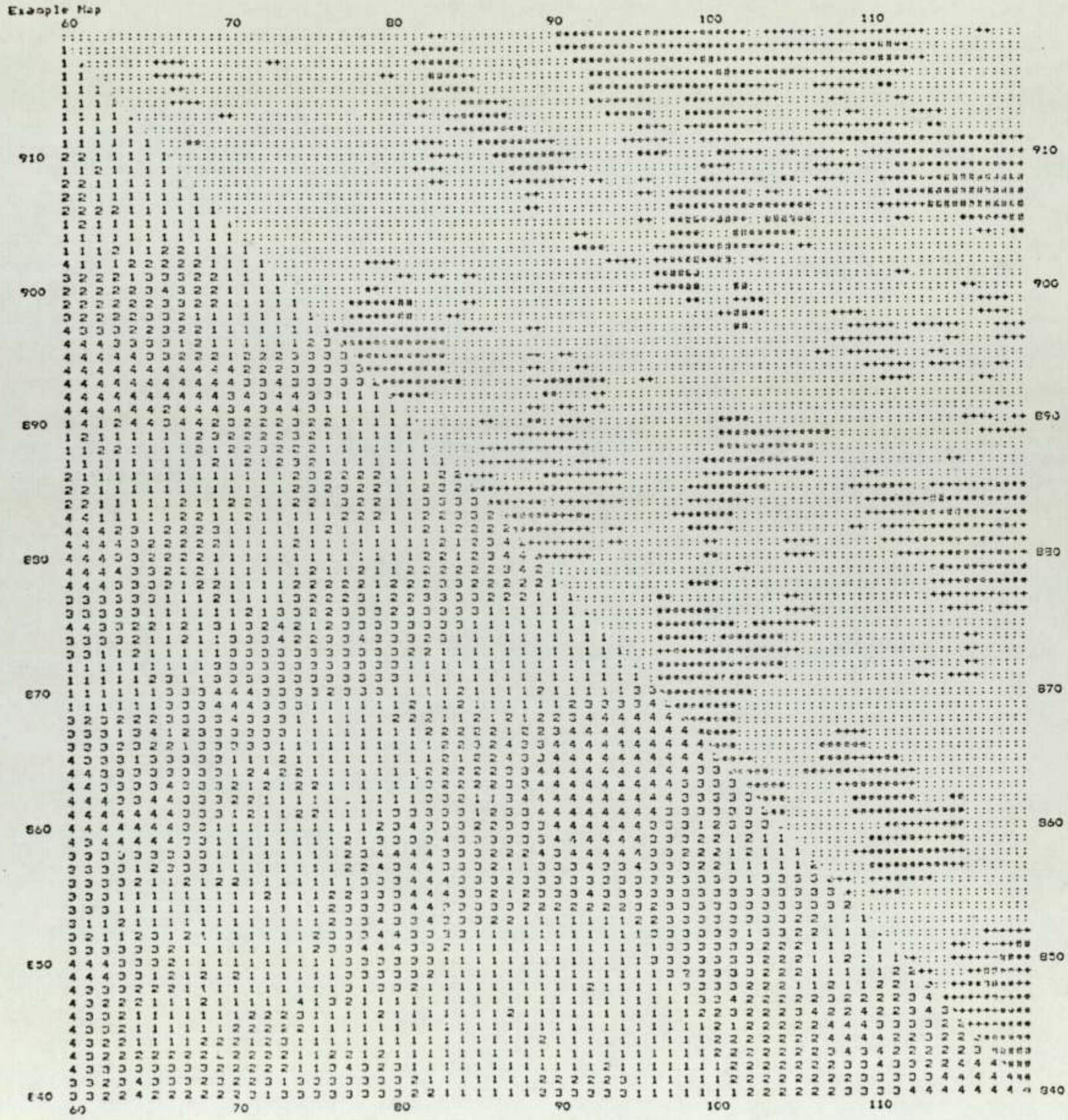
5.2.1 Isopleth maps

In this case the variable is an integer or a scalar value which has been divided into a number of classes. The number of classes that can be shown in shades of one colour is limited because of a combination of the eye's inability to reliably resolve between more than about seven shades on paper. This class division implicitly limits the 'vertical' resolution but the spatial distribution of the classes may be defined precisely. Examples of isopleth maps are the Geological Survey's maps of the United Kingdom or population density maps.

In recent years the computer has been used to produce isopleth maps on the line printer, Coppock in Davis and

McCullagh (1975). These maps have the disadvantage of a limited resolution in the X and Y dimensions but are cheap and easy to produce. They provide a 'quick look' capacity to mapping systems. Figure 5.1 is from CHORO, the choropleth mapper written for this project.

Figure 5.1 Sample output from the choropleth mapper



```
NUMBER OF CLASSES  5 CLASS BOUNDARIES
 83.50 2270.00 4452.00 7072.00 9787.00
SYMBOLS USED FOR EACH CLASS
    .          1          2          3          4
```

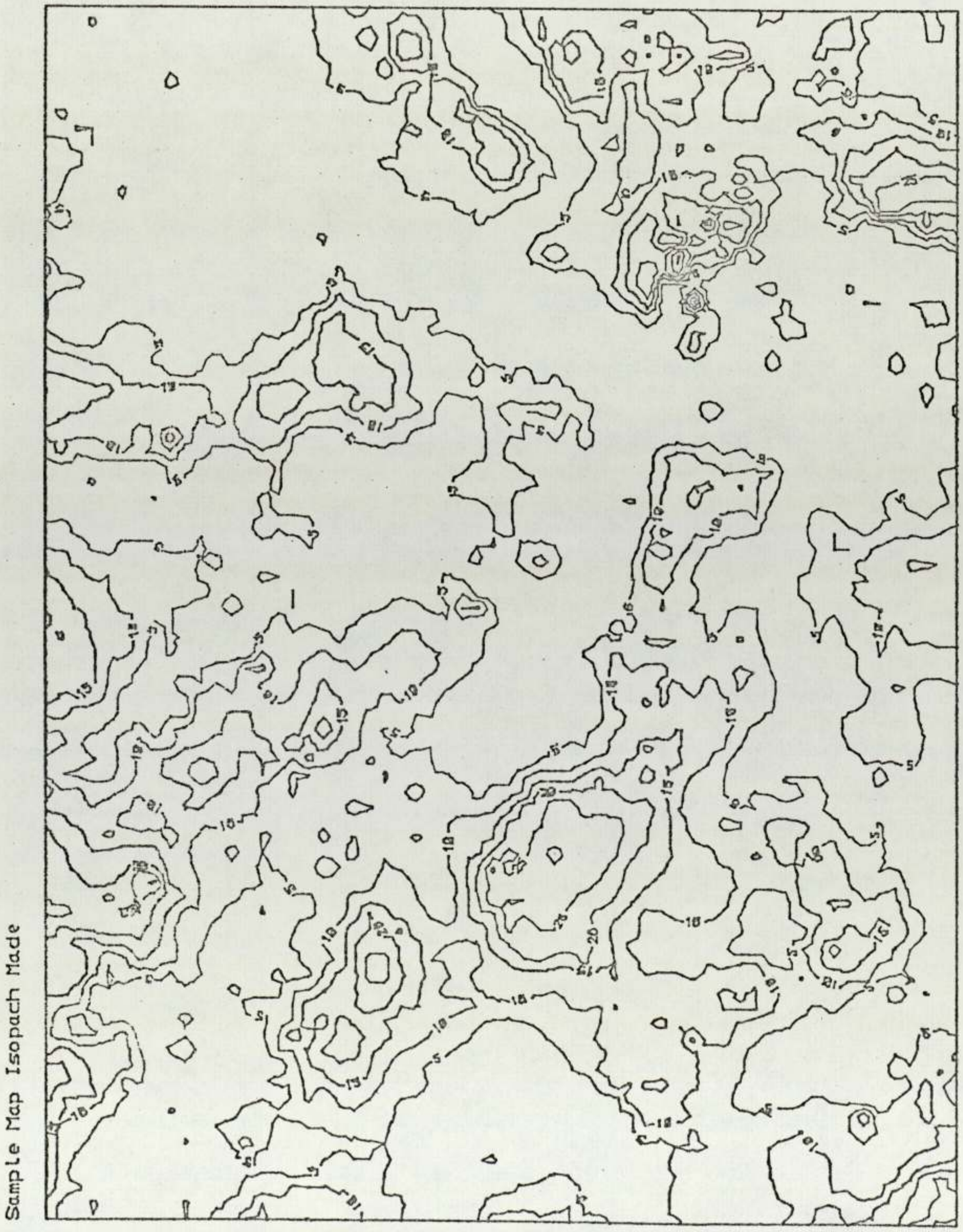
5.2.2 Isoline maps

These define real surfaces by contouring. The map is accurate at the contours but, of course, has to be guessed between them. Measurement of the surface tends to be easy but seeing a 'picture' of the features is more difficult. When an isoline map is drawn by hand the resultant surface is based on real data, qualifying general knowledge and preconceptions of the appearance of the final map. The computer has no world knowledge outside its programming but hopefully it is free from preconceptions of surface shape. This leads to some problems of reality but the map is reproducible and true to the original data - however good or bad that is. To draw smooth contours the program has to use a curve fitting routine. This can sometimes colour the results but rarely leads to great inaccuracies. The example, Figure 5.2, is plotted by GINOSURF and shows a well behaved plot with only a few sets of concentric rings indicating the holes or peaks that can occur when one point is very different from its neigbours.

5.2.3 Hill shading

This technique artificially illuminates a surface and is often used to enhance contour maps of topography. It gives the appearance of an aerial view under ideal conditions. The digital computer can calculate the luminance map for a surface faster and more reliably than a human, by repeated Lambert's cosine law calculations on very small, flat elements of the stored DTM. The final image may be used independently of the contour map to give an impression rather than a quantified map of the surface. However, even using a computer good results require a very fine mesh on

Figure 5.2 An example of a computer produced isoline map



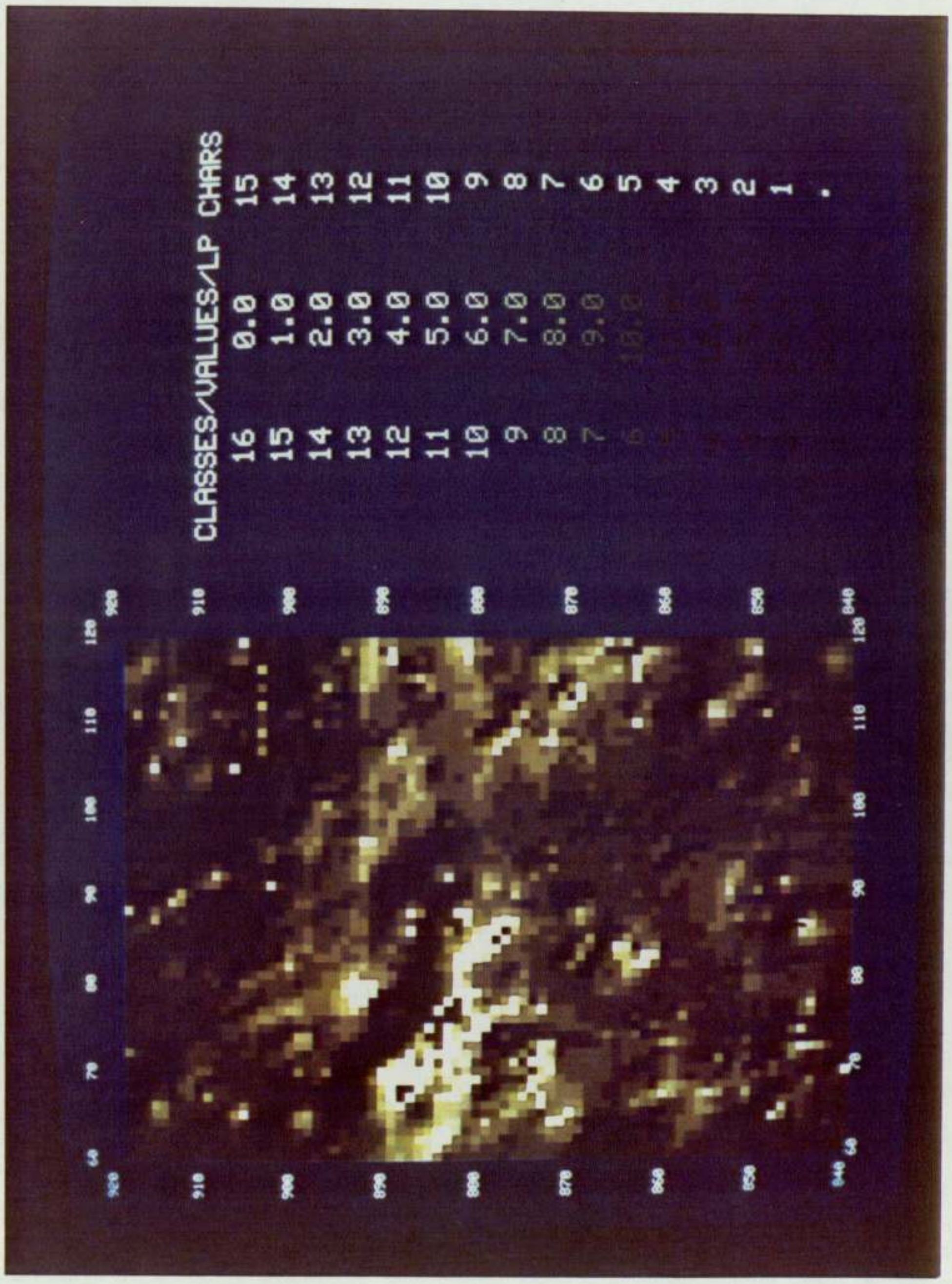
the luminance map and careful positioning of the 'light source'. The shading gives great subtlety to a map, enhancing the fine variations in slope which other methods do not resolve. The example, Plate 5.1, is plotted by CHORO and is of the rockhead of the mapped area, with the illumination in the northeast quadrant of the picture. For the best results the image should be viewed from a distance of at least two metres.

5.2.4 False three dimensional representation

There are many ways of viewing a geographic surface. The easiest to produce is the AXONOMETRIC view. This has a 1:1:1 axis relationship and therefore it can be measured directly with a ruler. Its appearance is not natural. The ISOMETRIC plot has its axes at 120° and can be measured by the use of a suitably scaled triagonal overlay. The viewing position of the observer is fixed to one of four points. Use of PERSPECTIVE gives more realism but the picture is only qualitative.

In all real three dimensional scenes some information is obscured by the concealment of portions of the image behind other parts of that image. To carry out this removal on the computer-produced picture, the computer being blind, requires another viewing operation. Several techniques are available for this process, they are known generically as Hidden Line Removal. Of interest for geographic surfaces are the Visibility Profile method and the Painter's Algorithm. Visibility profiling works from the front of the picture to the back. It calculates and omits the areas which are concealed from the viewer. The painter's algorithm draws a picture from the back to the front. The later material

Plate 5.1 Hill Shading



CLASSES	/VALUES	/LP CHARS
16	0.0	15
15	1.0	14
14	2.0	13
13	3.0	12
12	4.0	11
11	5.0	10
10	6.0	9
9	7.0	8
8	8.0	7
7	9.0	6
6	10.0	5
		4
		3
		2
		1
		.

obscures that drawn first only if it lies behind it from the current viewpoint. Possibly faster, more flexible and certainly more complex solutions exist but the simple case of a topographic surface is adequately handled by these two methods.

5.2.5 Device specialities

Graphics output devices are each suited to a different representation of mock 3-D. Paper and DVSTs require hidden line removal by visibility profiling to prevent cluttering of the image. The resultant image is of high definition and is suited to the production of anaglyphs for improved realism. Refresh displays give images which may be moved or rotated on the screen in real time to elicit their spatial sense which makes hidden line removal less necessary. The size of the area is limited by the number of vectors which can be plotted but the exciting possibility of interactive alteration of the image in real time becomes possible. Raster scan allows coloured, solid pictures to be built up. The painter's algorithm can be applied without any extra programming because this machine only retains the last change to the image. But it should be noted that the image cannot be drawn to matrix or inkjet printers without further processing. Textures and hill shaded 'tiles' can be applied to the displayed surface to give a realistic block model on the screen, or the whole screen can be made to appear like a photograph of the reconstructed topography.

5.2.6 Experiments in 3-D representation

For the current project a number of mock 3-D plotting routines were developed. Their case histories show that even after selection of the best plot and viewing routines, removal (or not) of hidden lines, choice of realistic or schematic representation and matching of method to machine problems can still occur....

The most correct presentation of the grid which is used to co-ordinate the data in the map database is as a surface made up of squares lying parallel to the X-Y plane. A program, DPLOT, was written to draw surfaces in this form, using GINO-F three dimensional plotting routines. These enable the picture to be viewed from any point relative to the X,Y,Z axes and projected on to the view plane, either axonometrically or in single point perspective. An added advantage of employing GINO is that the plot can be draw on any device. Anaglyphs such as Figure 5.3 & 5.3a may be produced simply by drawing the image twice with slightly different eye positions. The program has no hidden line removal making the image very complex. This renders the program unsuitable for temporary display purposes unless the image can be rotated to elicit more information about the structure. Only a refresh device would be fast enough for this application but none were available.

An attempt was made to produce a visibility profile routine which could give SYMVU (Fisher, 1968) style plots from any view angle. The complexity of the program made it too difficult and slow to use reliably. In an effort to provide a fast display for the system, and display as much information as possible, the colour graphics terminal was used to plot painter's algorithm surfaces by superimposing axonometrically viewed X-Y profiles. The surfaces could be coloured from a choropleth map file or by a simple, detail enhancing, slope shading technique. This program overloads the capabilities of the video lookup table and communications

A view of the proto-Tame Valley

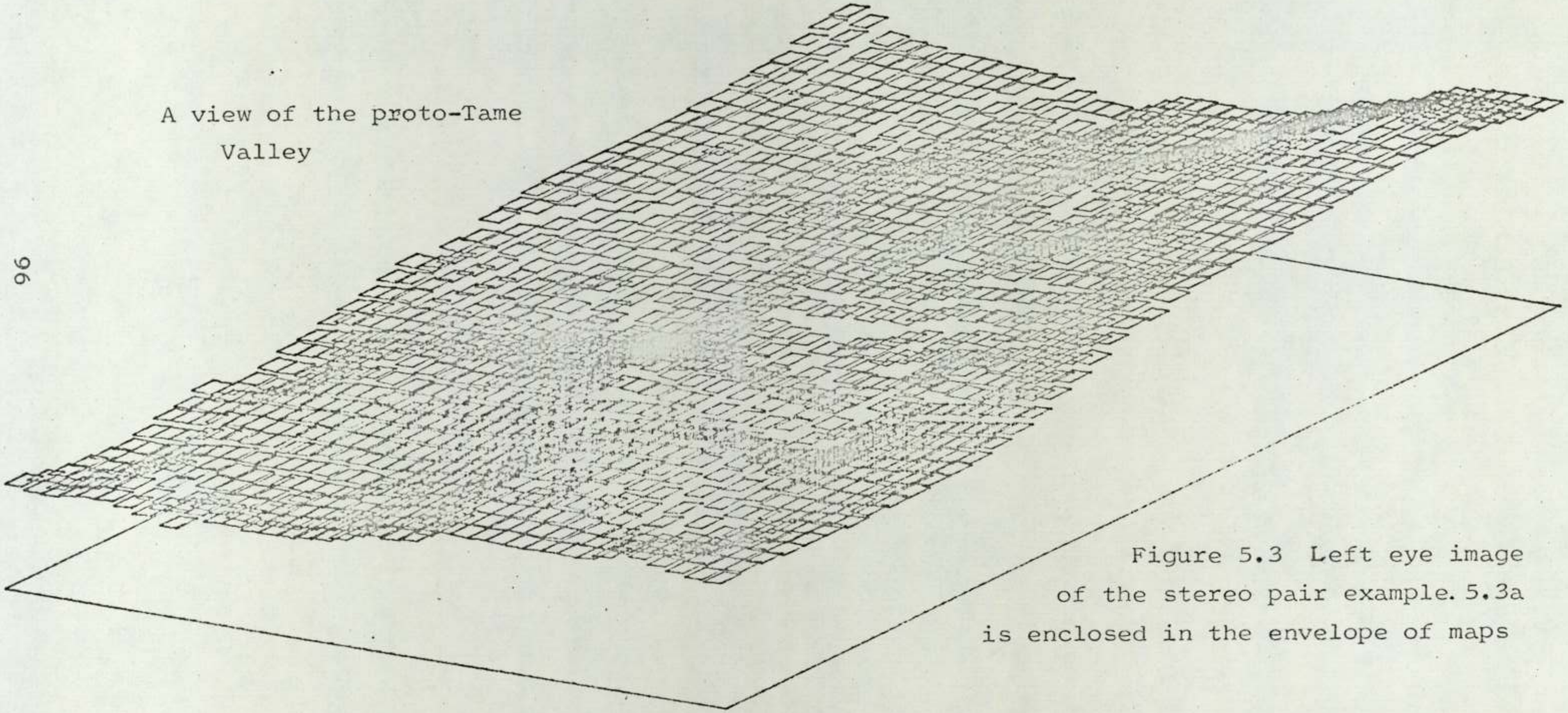


Figure 5.3 Left eye image of the stereo pair example. 5.3a is enclosed in the envelope of maps

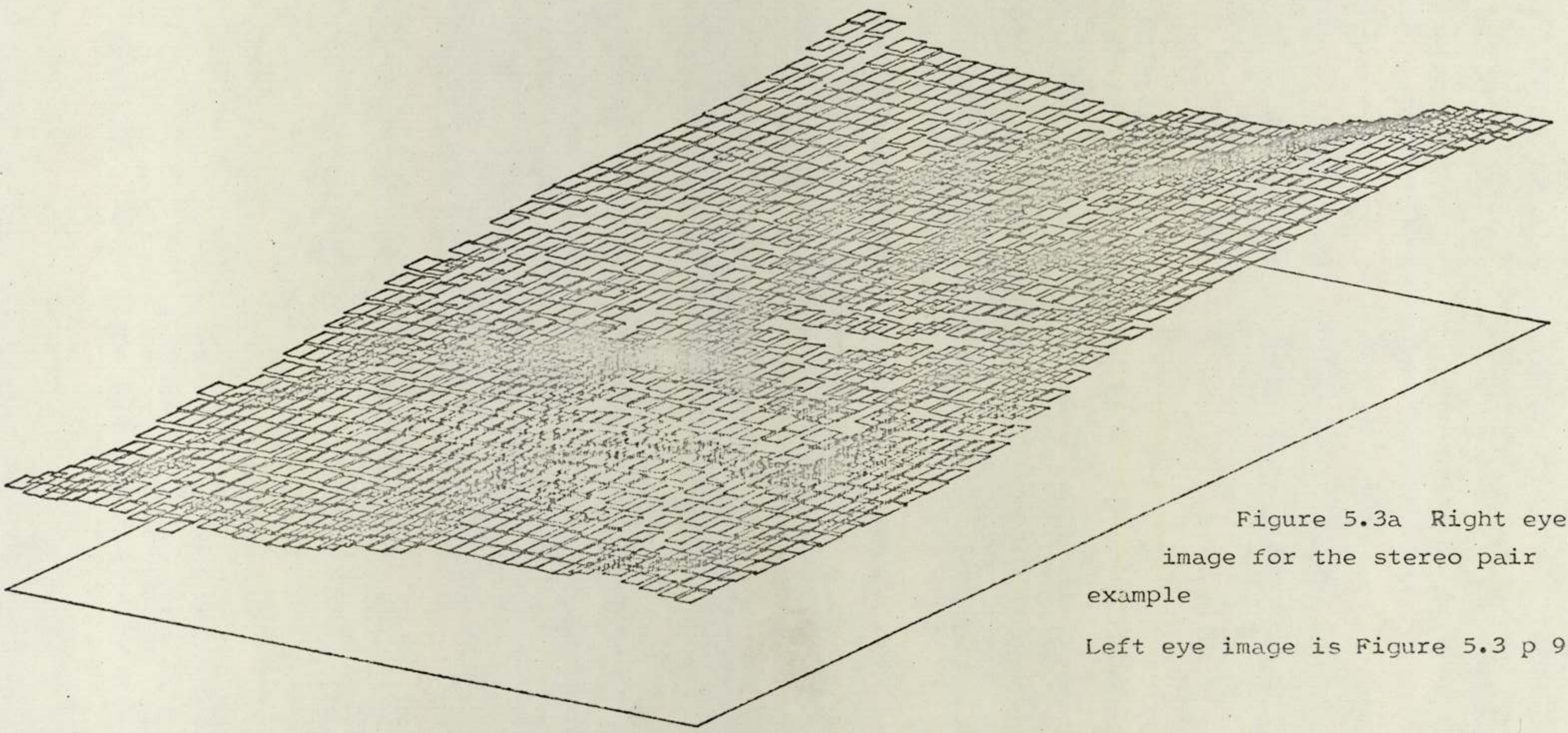


Figure 5.3a Right eye image for the stereo pair example

Left eye image is Figure 5.3 p 96.

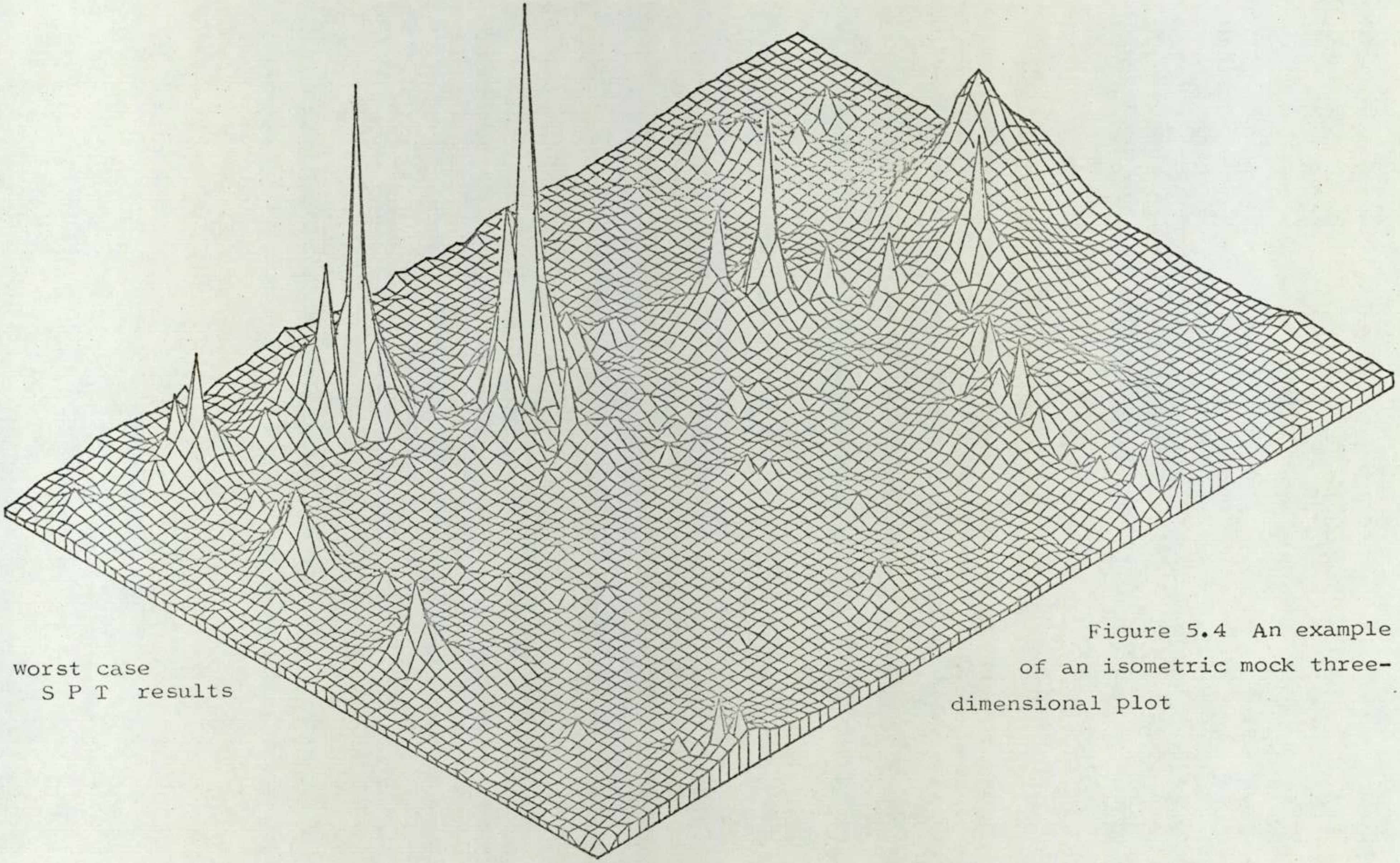


Figure 5.4 An example of an isometric mock three-dimensional plot

interface of the current colour graphics terminal.

Another program, to plot realistic hill-shaded block pictures as the ultimate in surface representation for the map database system, was beyond the software capabilities of the graphics display's local intelligence. It also took an estimated ten minutes to produce one image, due to the limited line speed to the device.

Fortunately GINOSURF, a commercially produced isometric and contour plotter, exists in the GINO library. This has usable routines for line drawn images on all devices but without any of the specific enhancements, except hidden line removal, mentioned above. For reliability only GINOSURF is employed for the output of the final database system. The example shown in Figure 5.4 illustrates the advantages of the mock 3-D representation in graphically demonstrating the major features of surfaces.

5.3 Representation Of Colour

There are three types of systematic colour definition: verbal description, parametric description and tristimulus values. The first is the most understandable to humans and the last are the intensity levels used by, for example, the television tube's electron guns. In between are artist's and colour scientist's attempts to quantify colour.

5.3.1 Specification of colour

5.3.1.1 verbal description

In geology the commonest standard colour description is as recommended by the Quarterly Journal of Engineering Geology Working Party Report (1972)* used as follows:

* Referred to from hereonin as WPR'72

column 1	column 2	column 3
light	pinkish	pink
dark	reddish	red
	yellowish	yellow
	brownish	brown
	olive	olive
	greenish	green
	bluish	blue
		white
	greyish	grey
		black

A choice of a colour from column 3, qualified by values from columns 1 and 2, is used to subjectively specify the shade of the sample. This system is simple, understandable and easy to directly translate into numeric codes for computer storage. It has recently been given a new lease of life by Berk *et al* (1982) as a colour naming method for use in computer graphics called CNS. This verbal technique has been reverted to, in a slightly more complex form than the three columns, because of the difficulties of visualising colours in the other, purely numeric, descriptive systems.

5.3.1.2 parametric estimation

These are spectrum based classifications. The values are usually derived from comparison with a standard chart such as the Geological Society of America's rock colour chart (1970). Two systems exist:

HUE SATURATION INTENSITY

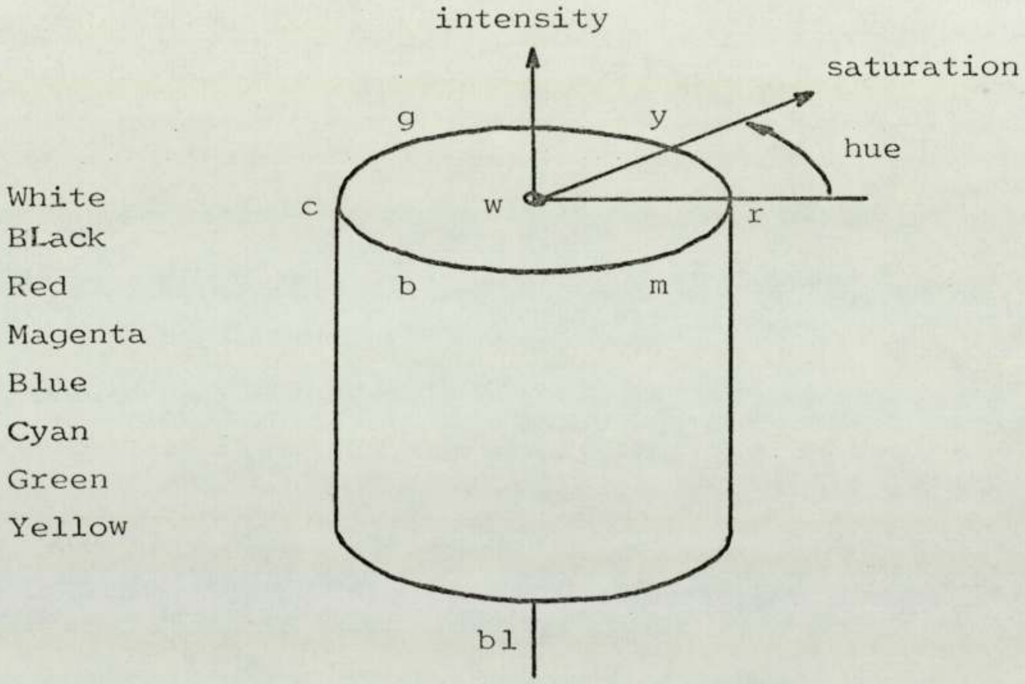
HUE classifies colour as red, yellow, green etc., i.e. the wavelength of the colour.

SATURATION, or chroma, determines the degree to which a colour differs from a grey or white of the same intensity.

INTENSITY determines how light or dark a colour is along a linear scale from black to full brightness.

Figure 5.5 shows the HSI cylinder. Hue is measured as an angle, Saturation as a distance from the cylinder's axis and Intensity as a position along the length of the cylinder. The base of the solid is black and the top contains the maximum intensity colours for different levels of saturation.

Figure 5.5 Hue Saturation Intensity

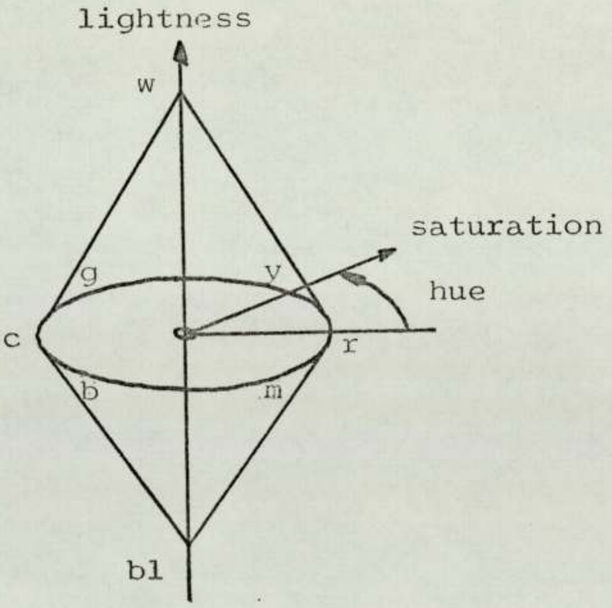


HUE SATURATION LIGHTNESS

This model, shown in Figure 5.6, uses a bi-cone to locate the colours (as Munsell, 1941). Here, maximum saturation is achieved with medium LIGHTNESS colours. The top point of the solid is white and the bottom black.

These techniques produce numerical values which can be converted into display compatible values without too much difficulty. HSI is easier to use and is finding favour with workers in the computer field but WPR'72 recommends the older HSL colour solid. Both methods of colour classification require a colour chart, though nowadays this could be an animated colour graphics display rather than a piece of paper. Visualisation of the colours specified also requires a chart for accuracy but a guess is usually close to the original specification. Close, that is, for an individual acquainted with the parameters - to anyone else the colour is unreadable. As has been mentioned before, translation to a verbal form or visual display is always necessary with a parametric encoding.

Figure 5.6 Hue Saturation Lightness

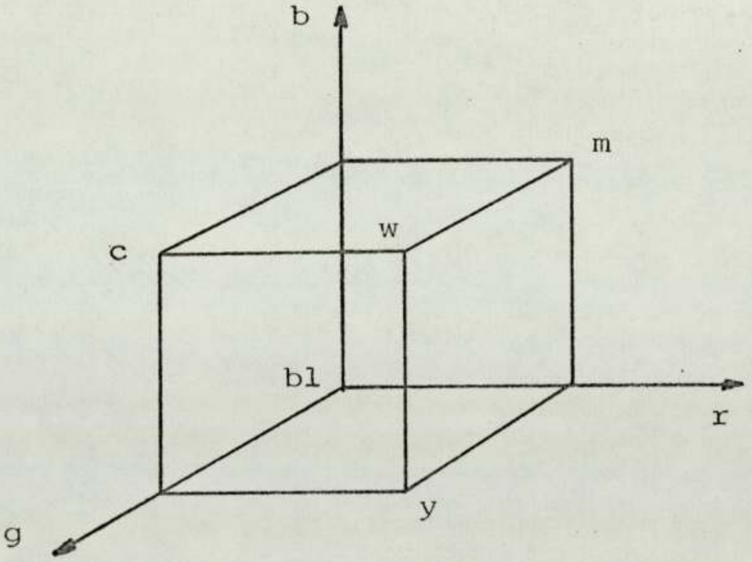


5.3.1.3 tristimulus values

These are the proportions of the primary colours in a particular colour. In an additive system (television, film recorders) the primary colours are Red, Green and Blue (RGB). For a subtractive system (colour matrix printers, inkjet plotter) Yellow, Cyan and Magenta (YCM) are used. Figure 5.7 shows the RGB colour space visualised as a cube in which the principal axes are the proportions of red, green and blue respectively. The grey scale lies from 0,0,0 which is black to white at 1,1,1. All possible colours lie within the cube. The solid is easy to imagine but the position of the colours within it are not.

On a colour graphics display the manipulation of the cube becomes easier, especially if a browsing program such as THREE written for and provided with the database system is employed. RGB is not compatible with sample nomenclature but whatever other naming system is used will have to be directly or subjectively translatable to this representation.

Figure 5.7 Red Green Blue tristimulus



5.3.2 Use of colour

Colour in graphical displays is useful for coding information, highlighting important detail and producing natural images. In geology, despite its often misleading interpretation, the colour is almost always recorded for a sample and can be reproduced for realistic representation as well as the more diagrammatic uses of general display enhancement.

The DVST, refresh devices and, to a lesser extent, X-Y plotters are often only monochrome, if colours are provided they are limited in number and are used for lines, not solids, with a black or white background. Realism, by shading and background colour washes, may be achieved through the many colours available with raster displays or matrix/inkjet plotters. Although, unless a lot of money is spent, the resolution of the image is not as fine as that produced with the line drawing devices.

Specification of the colour to be used will often depend on the device in use. High level graphics libraries make this specification easier but tend to limit the choice. The colours on a raster display, being determined by the numbers in the VLT array can be dynamically specified by a program. This allows great choice and flexibility making colour an interactive tool in the transfer of information.

The examples shown are of the RGB cube colour selection program THREE, Plate 5.2, and a set of colours produced using the numeric decoder for the database, Plate 5.3. Numeric codes were employed for colour naming for the reasons given in the preceding chapter. Translation to RGB is via a subroutine, COLOUR, which calculates intensities from a subjectively determined set of primary and secondary RGB values held in an internal register. A list of the subjective colour numeric codes is given in Appendix 2.

5.4 Representation Of Texture

Textures may be added to an image by shading or the addition of repeated character symbols, stippling. The method employed depends on the time which can be allocated to drawing the image and the device used.

5.4.1 Device-suitable textures

X-Y plotters and the DVST are line drawing devices which have virtually no limit to the amount of information that

Plate 5.2 Animated RGB colour solid THREE

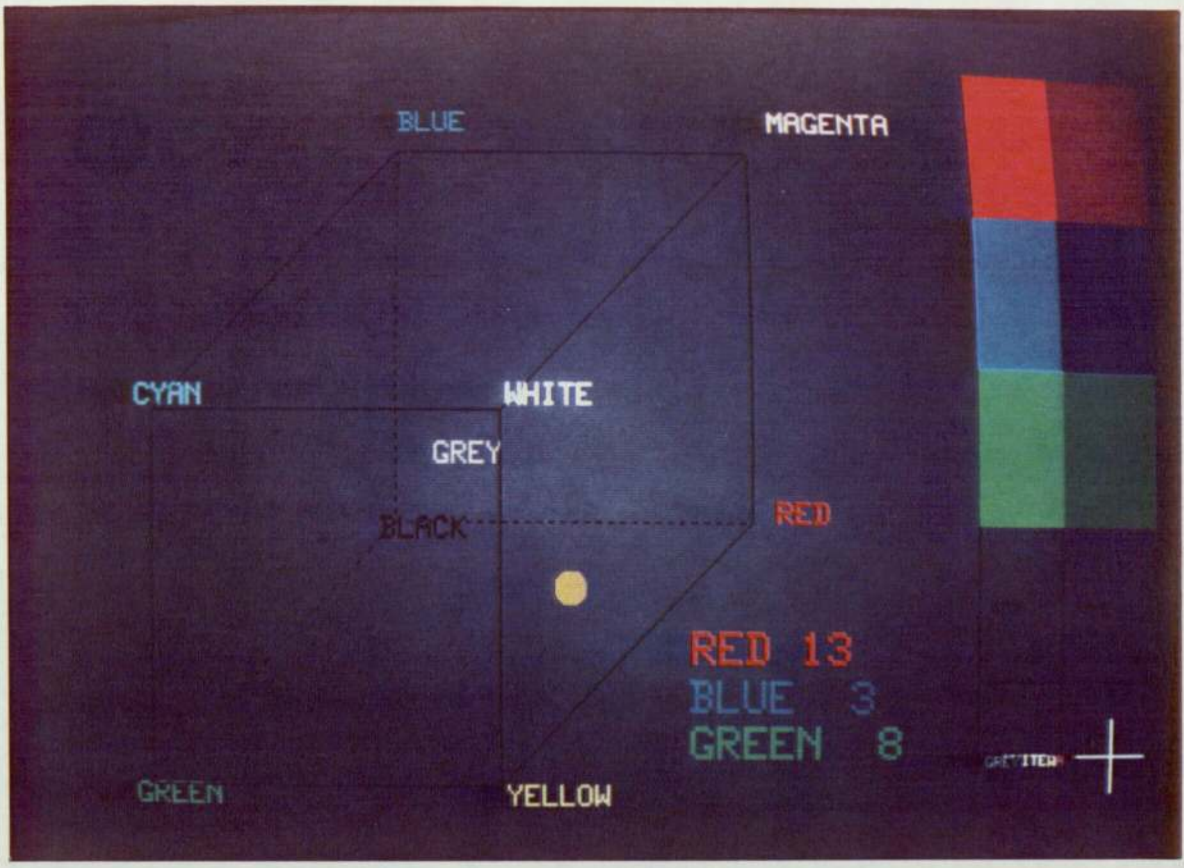
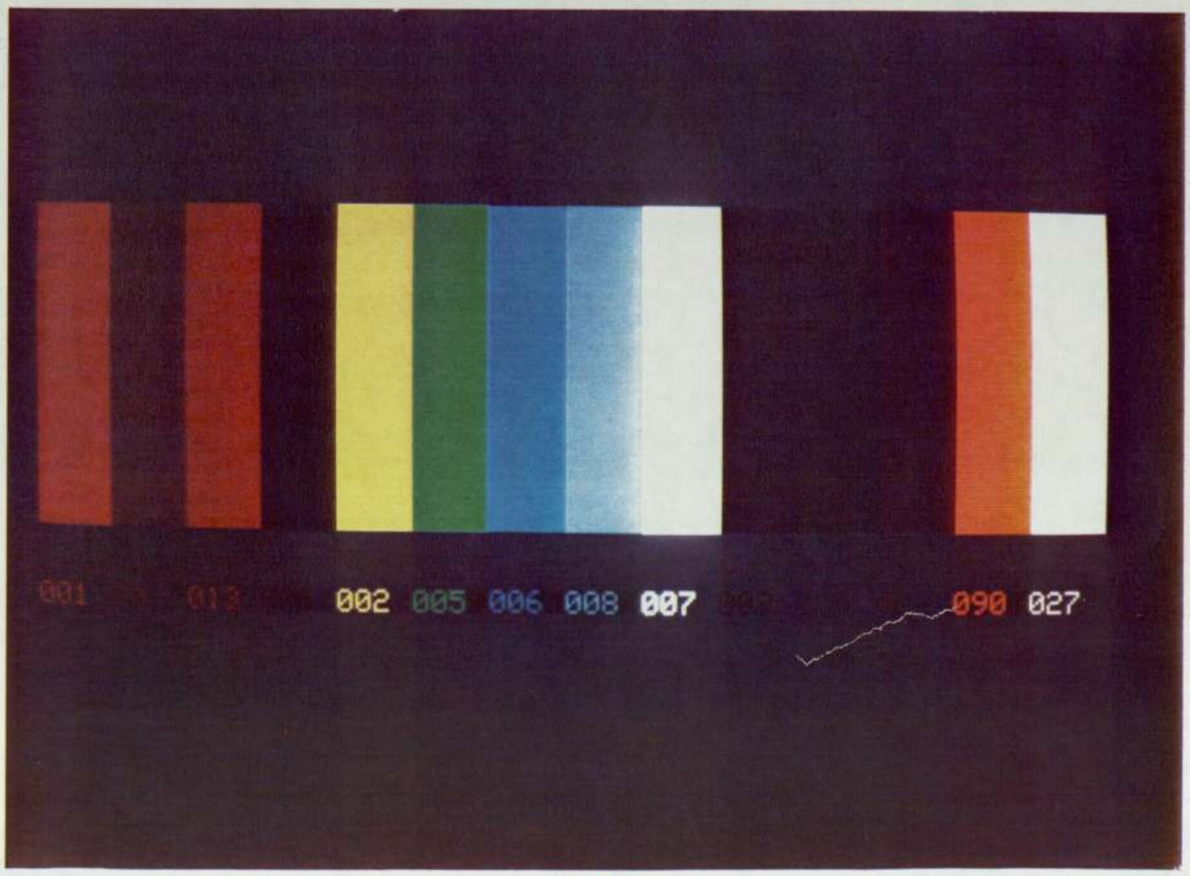


Plate 5.3 Sample colours from MUNSELL



can be displayed, but take some time to produce the image because only one pen is in operation. In this case the fastest texturing is shading because the number of pen movements are minimised. Several algorithms are available for the calculation of the shading of polygons, with both continuous and broken lines.

Refresh displays have the limitation that the total length of line which can be drawn on the screen is limited. This implies that the amount of texturing is limited to coarse shading or a few simple symbols.

Raster displays can cope with vectors but work much faster with symbols. These are held in a CHARACTER STORE. The symbol's 'box' can be positioned anywhere on the screen and the stored elements mapped onto the screen buffer, or BIT PLANES. The symbols can be designed by the user. This gives great flexibility to the textures employed. Advanced display processors allow for polygons to be filled with texture as well as colour. This makes the production of good high resolution maps or sections very simple indeed.

Matrix printers offer the same flexibility as the raster display, if suitable software is provided. However, if the device is connected as a line printer, or if only a line printer or alphanumeric VDU is available, the position of the character box is restricted to the vertical and horizontal character spacings. Shading is not satisfactory on the line printer but using stippling and symbolic texturing the line printer is employed extensively in the production of pictures and isopleth maps. This is not only because of the low cost of the output but also because of the ability to quickly fill areas with different densities of ink, rather than the accurate but spidery line drawings.

5.4.2 Examples

The following examples show texture symbols which have been employed in the map database set of programs. Plate 5.4 gives a graphical representation of the types in the list of lithology codes. The textures are calculated from the approximate proportions of the major grain sizes or special constituents, such as peat, in the lithological type specified. On the graphics display these symbols follow the WPR'72 recommendations, as far as is possible within the limitations of the device. The paper copy facility gives the WPR'72 alternative alphanumeric characters. The names for the numeric equivalents are given in Appendix 2.

Figure 5.8 shows the types of shading provided in the DRAW drafting program, which is designed for X-Y plotter output. The symbol approach is used because of the format of the map array files. Writing for pen efficency is essential to give a reasonable drawing time but despite care with this parameter the best time achieved whilst plotting a map which covered the whole area was over 20 minutes. This illustrates the superiority of the matrix printer in producing shaded maps. CHORO takes rather less than two seconds to produce a similar, slightly less accurate, map.

Figure 5.8 Texture symbols for the DRAW program

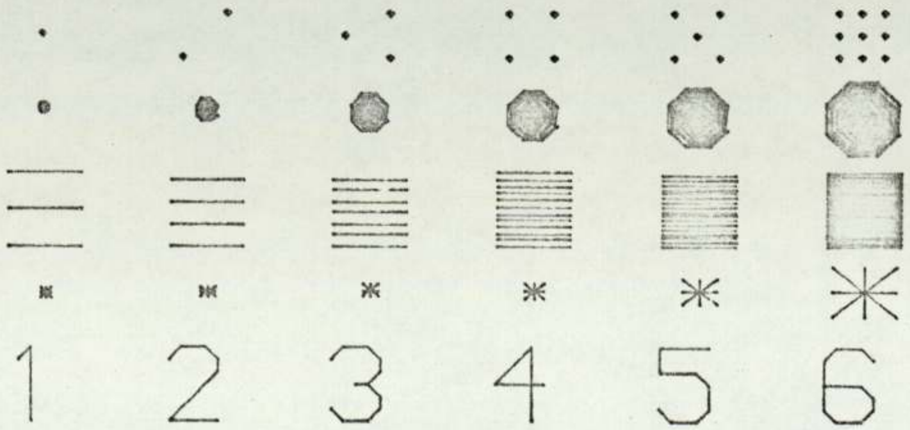


Plate 5.4 Graphics terminal synthetic textures

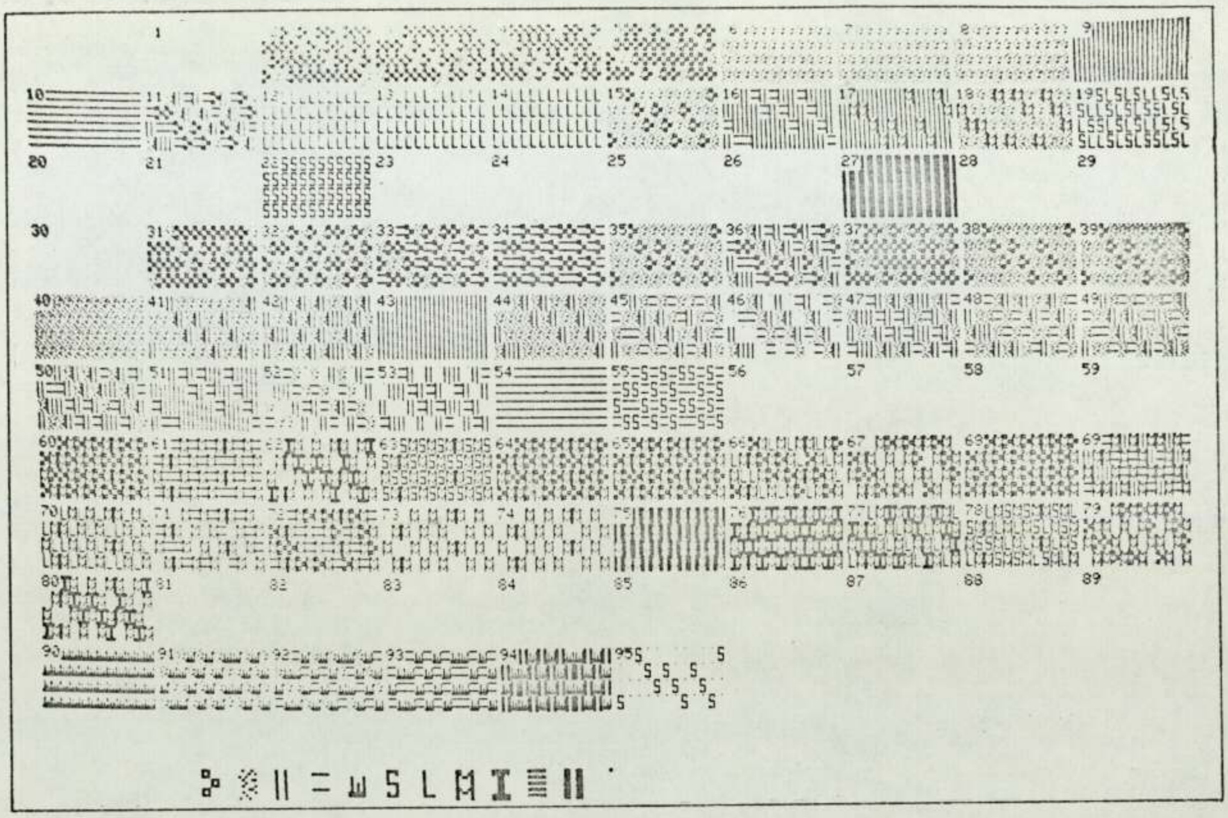
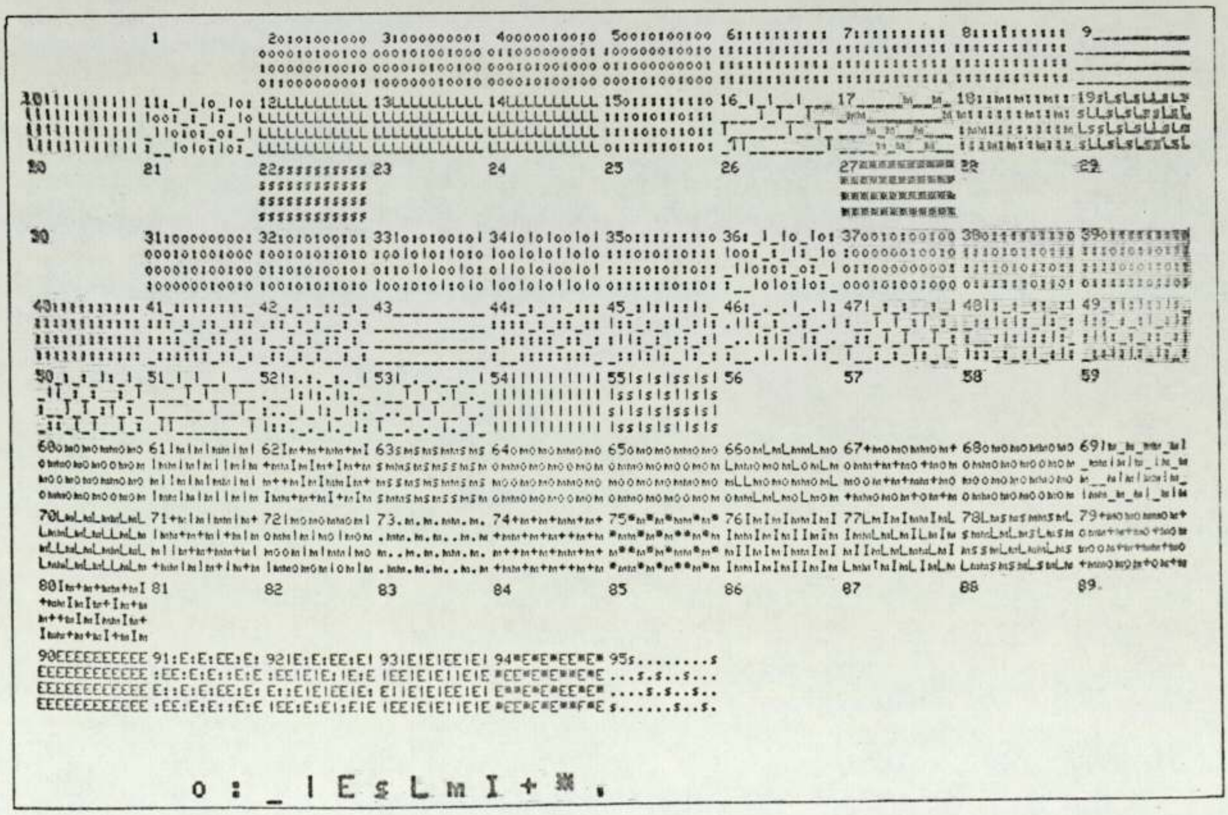


Plate 5.4a Line printer synthetic textures



5.5 Image Interaction

5.5.1 Workstation concept

An important concept in mainframe computer graphics is that of the WORKSTATION. A workstation is a set of graphical display devices with a collection of associated input tools designed to work together in an integrated manner. These physical devices are 'animated' by a master program. This employs graphical control subroutines for co-ordinating input and output and is called the USER INTERFACE. Any user interface may be divided into four linked components:

The USER'S MODEL. A conceptual description of the function of the program or system which allows an individual with little or no knowledge of computer technology to understand its operation. An example of a user's model is the datastore structure of this project as described in Chapter 7.

COMMANDS. Ideally a command language should be natural, facilitating immediate understanding of the model as a device capable of performing a set of operations. Examples are the typewriter or calculator which may be used with only general knowledge as prior experience.

FEEDBACK. Assistance of the user by the computer. This comes in many forms: acknowledgment of receipt of commands, explanatory messages, indication of selected objects, cursors, echoing of typed characters, inking in of drawn lines, rubber banding the cursor to the previously drawn point. Some forms of feedback are provided mainly to help inexperienced users and may be ignored by experts. Others, such as the position of a cursor on a screen, are essential to the interaction of man and machine.

INFORMATION DISPLAY. This provides the user with the state of the information which he is manipulating. The image must be organised to convey these data as efficiently as possible and must be in accordance with the user's model. If the model is based on a reality it must be as realistic as possible. When a synthetic model is employed it should use well chosen symbols and graphic imagery.

5.5.2 Graphics tools

For interactive input to programs there are two classes of tool.

EVENT DRIVEN. These require action from the operator to send some information to the program:

choice - The return of a single value such as given by a function switch or keyboard.

identifier - A picture segment name from a light pen hit or a cursor driving device such as a joystick or a tablet and a pointing algorithm.

text - A string from a keyboard or handwriting recogniser.

timer - A programmed stimulus which causes an event, for instance, the updating of the display.

SAMPLED. Continuously read by the software or the local device processor and used when requested by the program:

valuator - a potentiometer or multiposition switch which can be set and left.

locator - Returns coordinates, for instance, a mouse or sensitive screen or light pen with a tracking algorithm.

Graphical tools can also be divided into three types depending upon the kind of data they return most effectively:

POSITIONING. Definition of either the location of points which can be used to construct lines, curves, polygons etc. or the positions of symbols and picture segments/parts. This is best done by locators, such as the trackerball or tablet, but may also be accomplished with a light pen and a tracking algorithm.

POINTING. Indicates to the program the part of the image which is of interest or identifies a segment. Pointing is carried out either by a pointing device such as the light pen or with a locator and a coordinate search algorithm. Both methods require a stored picture definition which can be searched for the indicated segment.

PARAMETER SPECIFICATION. This can be carried out by typed commands from the keyboard, readings of a valuator or by calculating the value of a locator on a displayed scale.

5.5.3 Use of tools

Tools are employed by the user interface to give input, output and selection of modes of operation. A number of techniques are available; which one is selected depends on the user's model, type of operation and, of course, the tools available at the time. Some uses of the various tools are as follows:

FUNCTION SWITCHES. Either as a choice tool or as a valuator. These are easy to use and program, generally available and fast for the user. But a complicated program will require a large number of switches or confusing re-definitions and, in addition, valuator use requires very fast communication with the computer.

TYPED COMMANDS. For sophisticated commands, text may require a complex decoder. These are not trivial to write and must be friendly to the user. Exact, recorded values can be given but the speed of interaction is usually slowed by the speed of typing, both physically and by remembering what codes to type. Extensive help and labelling facilities are required.

MENUS. These may either be displayed on the screen or on a card placed over a tablet. The required function is indicated by a pointing tool. Some feedback is needed to indicate which segment was 'hit' and what process is in operation. Menus require search software and, if on the screen, take up a lot of space. Hierarchical structures can reduce the display waste. Movable menus are ergonomic but require more programming.

CHARACTER RECOGNISERS. An advantage of the tablet input device is the ability to draw strokes which can be interpreted as commands. This allows a large number of functions to be used quickly but needs training to each user and requires some care in its use. Voice recognition comes into this class of tool. The software is sophisticated and the decoder requires training but in use this method is very fast.

5.5.4 Device/tool matching

Each display device has an input tool which is particularly appropriate to the type of graphical output produced by that device. The DVST with its pen and ink style is suited to the joystick/trackerball and an associated low intensity cursor cross which does not leave a permanent screen image. Refresh displays work well with the light pen. Movable menus, drawn to follow the pen, and the facility to drag pen-sensitised, highlighted segments around the screen allow this device to give full, fast interaction. Raster images are limited in resolution and menus use up valuable space. The tablet is ideal for sketching pictures, locating objects and can also issue stroke-defined commands not directly related to the structure of the screen image.

5.5.5 Examples of interaction

Two examples of image interaction are given. One, Figure 5.9 and Plate 5.5, is the menu driven data reduction MERGE program, showing the user and program operations for a semi real model. This displays dynamic data, uses positioning, pointing and parameter specification and feeds back on several levels. The other example, Plate 5.6, shows the typed command driven drafting program DRAW. In operation the program must operate from simple input devices and therefore cannot rely on special graphical commands. In this case the provision of examples and descriptions becomes important, leading to a different program structure, Figure 4.12.

5.6 Hard Copy

5.6.1 High quality drafting

If a large number of users are to be catered for the ability to obtain permanent copies of information from database enquiry sessions is a necessary part of the system. The plotters provided at Aston have already been described. These are employed for high quality output via the drafting program DRAW. The graphics in this program are largely left to the GINOSURF package, the choropleth routine being written to emulate the contour routine options. A number of different shading types are made available, not only to provide flexibility but also to give effects suitable to the speed and resolution of the current device. Colours following the GINO standards are provided to give effect to isopleth maps and contrast to overlaid isolines.

5.6.2 Fast graphics output

Line printer output is relied upon because of its speed

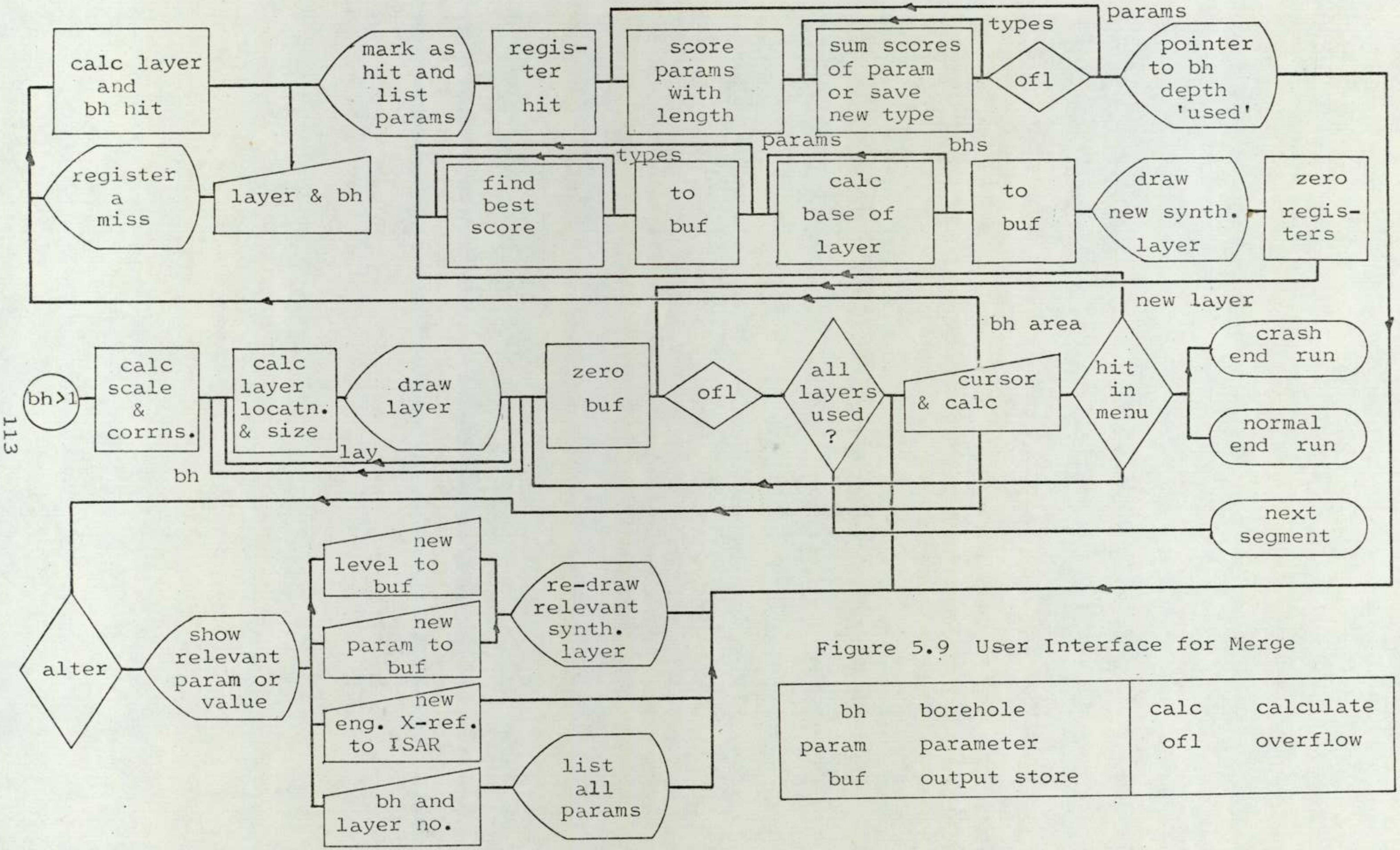
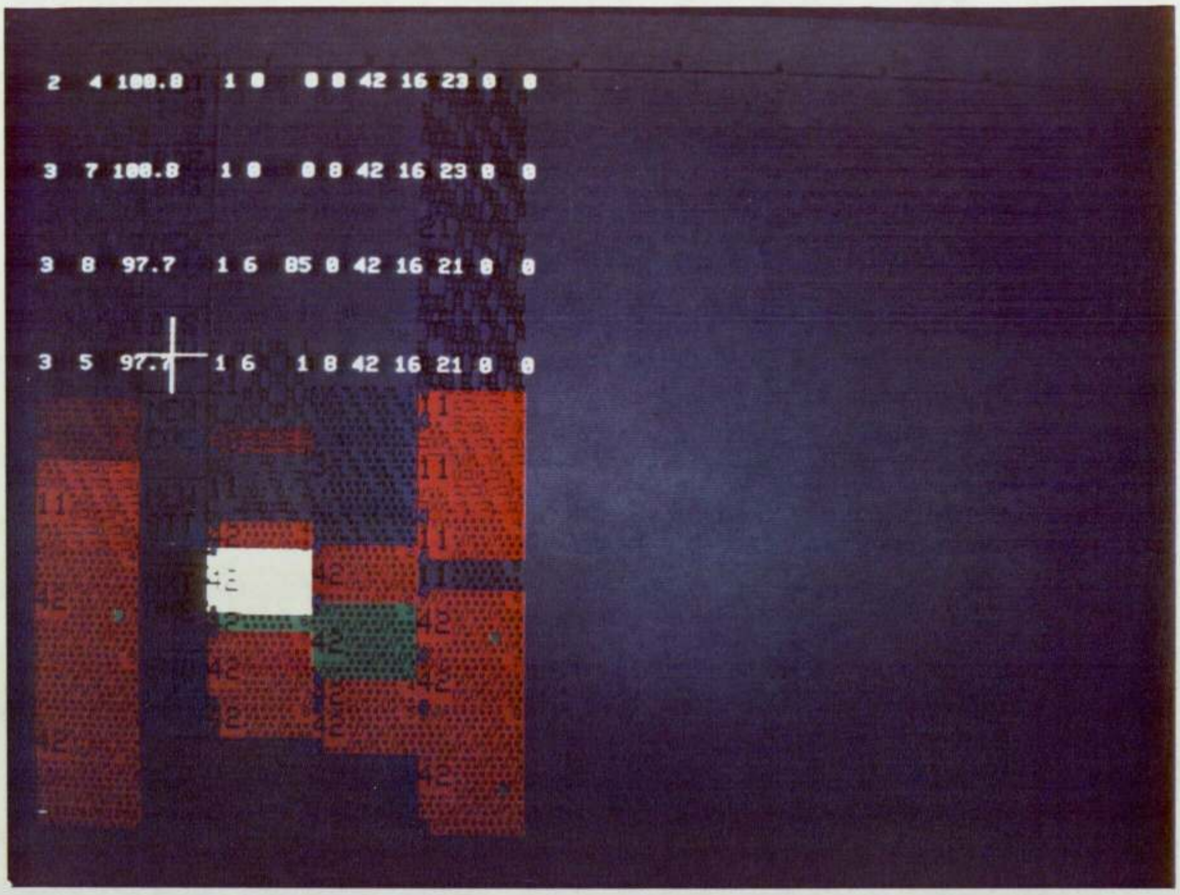


Figure 5.9 User Interface for Merge

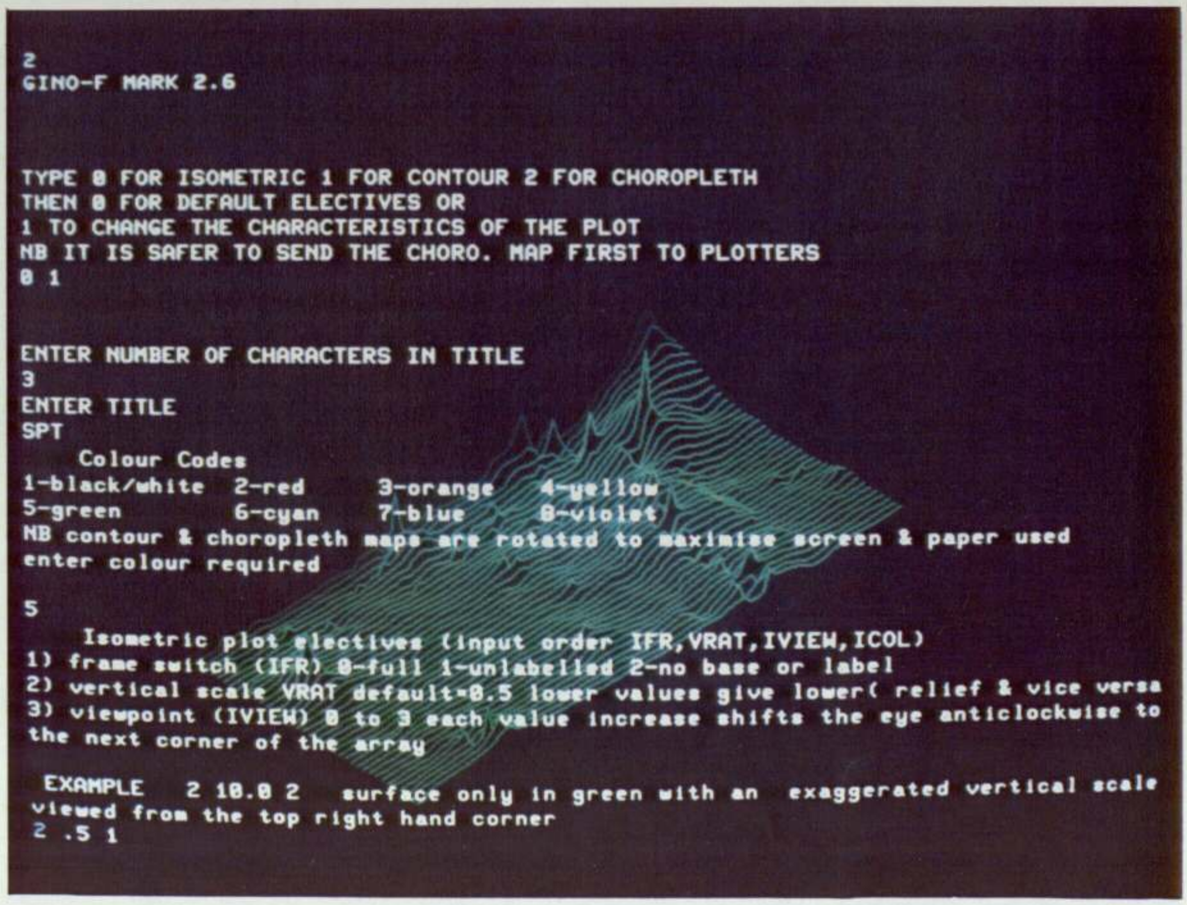
bh	borehole	calc	calculate
param	parameter	ofl	overflow
buf	output store		

Plate 5.5 Borehole synthesis MERGE



```
2  4 100.0   1 0   0 0 42 16 23 0  0

3  7 100.0   1 0   0 8 42 16 23 8  0

3  8  97.7   1 6  85 0 42 16 21 0  0

3  5  97.7   1 6   1 8 42 16 21 8  0
```

Plate 5.6 Drafting with DRAW



```
2
GINO-F MARK 2.6


TYPE 0 FOR ISOMETRIC 1 FOR CONTOUR 2 FOR CHOROPLETH
THEN 0 FOR DEFAULT ELECTIVES OR
1 TO CHANGE THE CHARACTERISTICS OF THE PLOT
NB IT IS SAFER TO SEND THE CHORO. MAP FIRST TO PLOTTERS
0 1


ENTER NUMBER OF CHARACTERS IN TITLE
3
ENTER TITLE
SPT
    Colour Codes
1-black/white  2-red      3-orange    4-yellow
5-green        6-cyan     7-blue      8-violet
NB contour & choropleth maps are rotated to maximise screen & paper used
enter colour required

5
    Isometric plot electives (input order IFR,VRAT,IVIEW,ICOL)
1) frame switch (IFR) 0-full 1-unlabelled 2-no base or label
2) vertical scale VRAT default=0.5 lower values give lower( relief & vice versa
3) viewpoint (IVIEW) 0 to 3 each value increase shifts the eye anticlockwise to
the next corner of the array

 EXAMPLE   2 10.0 2   surface only in green with an  exaggerated vertical scale
viewed from the top right hand corner
2 .5 1
```

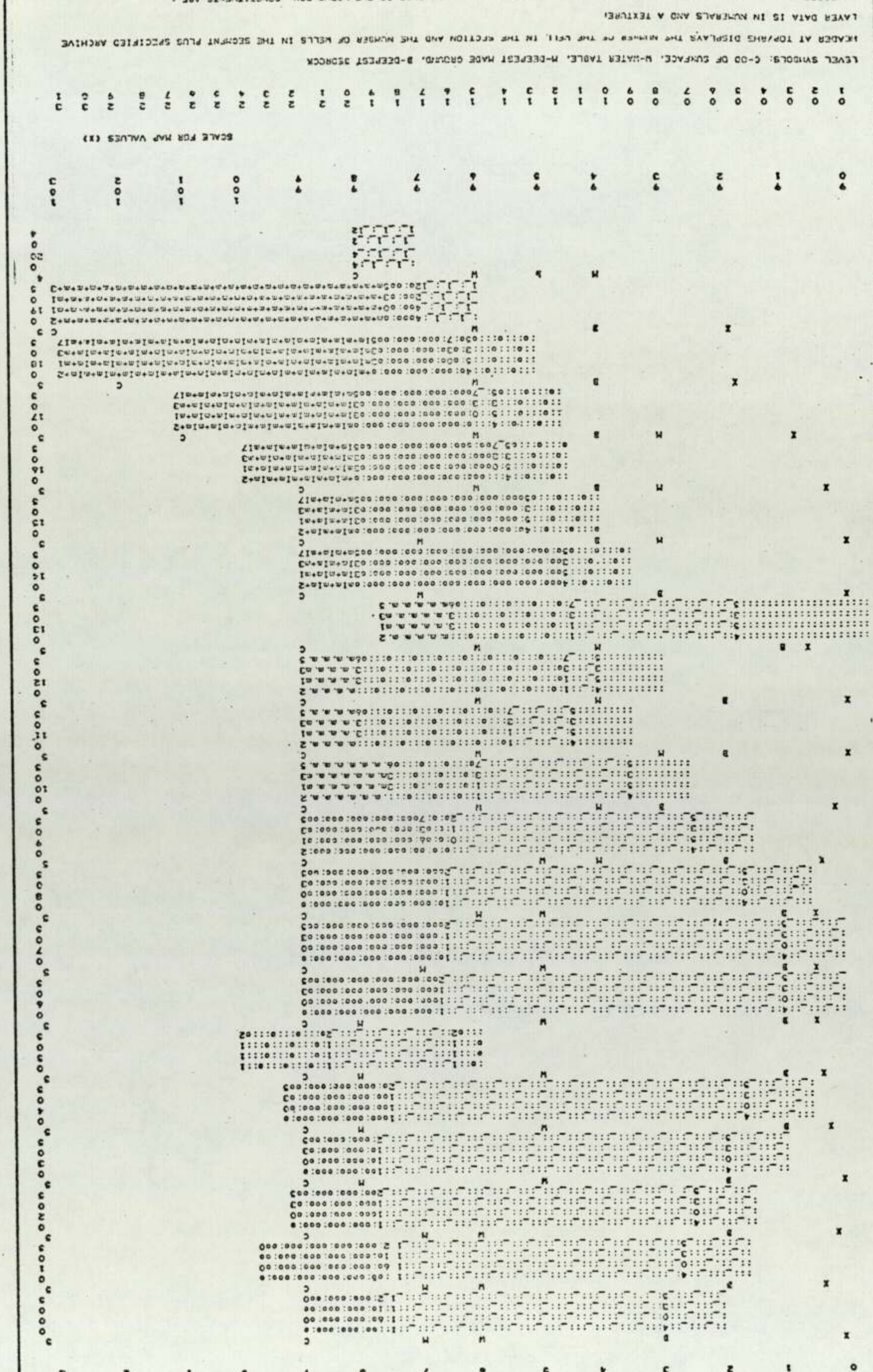
and low cost. The printer in most multiuser systems is not a resourceable device and therefore the programs have to write to an output file. This file is then listed to the printer at the user's request when the program terminates.

To produce graphical output on a line printer requires an approach similar to raster graphics display. An internal buffer is treated as a 'screen' image. The array has to be filled with space characters before other text is written to it. This text provides the elements of the image. Textures are applied by a loop which assigns values to the addresses to be shaded. Numbers have to be split up by successive remaindering into their discrete integers. These are added to the appropriate addresses in the buffer via a look-up list which provides the ASCII text code for that number. The buffer is written out row by row to the output file. Care has to be taken to print the data to the display file in the right order, because it is not possible to rewind the file or paper to add information. Text labels may be added around the buffered sections by using standard format commands.

The example given in Figure 5.10 is from the SECTION program and illustrates the WPR'72 text texture character set in use. The section is printed sideways, down the paper, to allow any length of section to be plotted. The vertical resolution is thus limited to 130 characters across the paper. The maximum length of borehole proved in the mapped area is 40m. This gives a minimum resolution of about a third of a metre per character, as compared with a resolution of a tenth of a metre in the four significant figure standard for the datastore.

The isopleth maps from CHORO, Figure 5.1, are another

Figure 5.10 Sample output from the cross section program



```
METRES                                                    W-E SECTION FROM SP 60919     N

SCALE FOR MAP VALUES (X)

LEVEL SYMBOLS: C-OD OF SURFACE. W-WATER TABLE. M-DEEPEST MADE GROUND. B-DEEPEST BEDROCK
HEADER AT TOP/RHS DISPLAYS THE NUMBER OF THE WELL IN THE SECTION AND THE NUMBER OF WELLS IN THE SEGMENT PLUS SPECIFIED ARCHIVE
LAYER DATA IS IN NUMERALS AND A TEXTURE:
THE FIRST NUMBER IS THE STRATIGRAPHY. THE SECOND IS THE STRENGTH. TEXTURE IS THE LITHOLOGY. CONSTITUENTS ARE :
   o GRAVEL : SAND _ SILT 1 CLAY E PEAT \ SANDSTONES,SANDS ETC L LIMESTONE m MADL,MARL I * *  INDUSTRIAL HAZARDS . BLANK
```

example of line printer output. In this case the main disadvantage of the use of line printer format is the difference in the X and Y scale, due to the rectangular character box. Line printer compatible pictures can, of course, be plotted on any alphanumeric device. This can facilitate bulk production of very cheap copies by simply re-listing the output file and provides efficient storage of data in microfiche form. Appendix 5 is such a set of microfiche listings.

5.6.3 Text output

Text-only output also has to be formatted for ease of understanding. Upper and lower case and variable fonts may be employed to divide information, messages and commands. The layout should employ the whole area of the screen or paper but there is an overflow hazard in the varying size of the display area, measured in characters, of different devices. The example, Figure 5.11, is a SEGMENT single borehole log in alphanumeric text for VDU or line printer listing. It uses a look-up translation routine to decode the geological information into the WPR'72 standard English terms. This sort of user-friendliness is the ultimate aim of good, usable, interactive computer interfaces.

Figure 5.11 Sample output from the Segment program

codes

```
Grid Ref 61840 OD 122.0 Water 114.5( 1) Bedrock 112.5( 1) Made 117.5( 69)HZ0022
Eng. data: pH/lay 8.2( 1) sulphate 0.098( 1) SPT   8( 3) max shear  0( 0)
strat  moisture content  bulk density      shear    strength  atterberg limits
              %                Mg/m**3             KN/m**2                 %
 0      192  1 192  2       2138  1 2138  2     35  1     35  2     31  1     11  2
 4      234  2 333  5       1959  2 1888  5     29  2     24  5     57  1     36  5
Layer data: number of wells  2 top of well @ 118.8 levels @ base of layer
117.5 1COL    0 Q 0 2COL    0 W 0 ST 21 LT 69 SN 13 Q 3 MN 30
115.8 1COL    3 Q 0 2COL    0 W 0 ST  3 LT 53 SN 13 Q 7 MN 27
114.7 1COL    3 Q 0 2COL    0 W 0 ST  3 LT 37 SN 13 Q 0 MN  0
113.8 1COL    0 Q 0 2COL    0 W 0 ST  3 LT 31 SN 35 Q 0 MN  0
111.1 1COL    3 Q 0 2COL    0 W 0 ST 42 LT 53 SN 37 Q 0 MN  0
```

translated

```
Grid Ref 61840 OD 122.0 Water 114.5( 1) Bedrock 112.5( 1) Made 117.5( 69)HZ0022
Eng. data: pH/lay 8.2( 1) sulphate 0.098( 1) SPT   8( 3) max shear  0( 0)
strat  moisture content  bulk density      shear    strength  atterberg limits
              %                Mg/m**3             KN/m**2                 %
 0     19.2  1  19.2  2  2.138  1 2.138  2     35  1    35  2    31  1    11  2
40     23.4  2  33.3  5  1.959  2 1.888  5     29  2    24  5    57  1    36  5
Layer data: number of wells  2 top of well @ 118.8 levels @ base of layer
117.5  MADE GROUND BUILDING WASTE LOOSE COMMON BRICKS & TIMBER
115.8  BROWN ALLUVIUM SILTY CLAY LOOSE WITH SAND & GRAVEL
114.7  BROWN ALLUVIUM SAND-CLAY-GRAVEL LOOSE
113.8  ALLUVIUM VERY COARSE GRAVEL VERY DENSE
111.1  BROWN KEUPER MARL SILTY CLAY MEDIUM DENSE
```

6 DATA ACQUISITION AND PROCESSING

This chapter describes the sources and quality of the information captured for the map database. The spatial and recording reliability of the information is analysed, with a view to establishing data processing techniques for the recorded values. The reduction of the data to the final datastore format is illustrated, to demonstrate the advantages and disadvantages of computer aided processing and artificial intelligence in borehole parameter handling. Figure 6.1 illustrates the main stages in data processing during the project.

6.1 Data Sources

6.1.1 Coordinate located

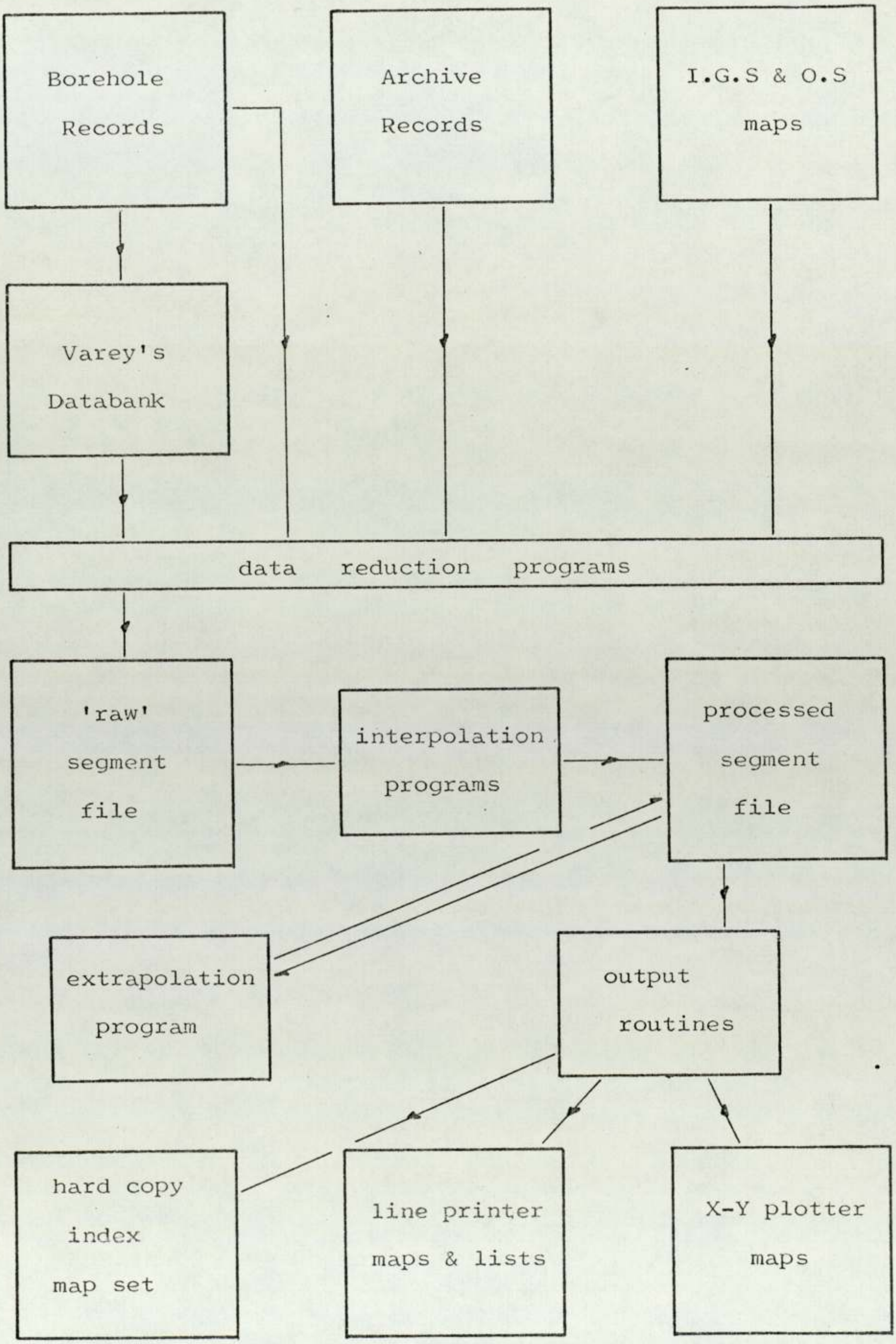
The spatial location of the collected information is defined in two ways, either by National Grid coordinates or implicitly - by virtue of its position on maps.

The coordinate located data used is from the site investigation reports held by the Birmingham City Council Structural Engineer's Department in Baskerville House and the registry of tips, pits and boreholes for dumping, maintained by the Severn Trent Water Authority at Alpha House. Further sources which may be used for extension of the area, or for more detail, are the Highways Department of the County Council (for tips and site investigations), local site investigation firms and the CIRIA recommended National Registry of ground investigation reports, when available.

6.1.2 Implicit location

'Implied' location data is collected on a grid. This

Figure 6.1 Stages in Data Processing



fixes the position of the information by its presentation and storage format, rather than by coordinate specification. The sources employed were the Institute of Geological Sciences (I.G.S.) solid and drift one inch to one mile sheets (for a reference stratigraphy), the Ordnance Survey (O.S.) 1:10,000 maps of the area and the City Council's archive maps for 1731 and 1825. In addition to these data a study could usefully be made of the O.S. Series One plans for the area.

6.1.3 Previous databank

Varey's (1977) databank holds coded boreholes on serial files with locations accurate to ten metres. In these files provision is made for the storage of all the parameters required by the geotechnical standards. Because the current project uses simplified codes and formats some degree of translation will be required before the data are usable. Varey's retrieval programs do not allow 'whole database' manipulations and some of his decoding techniques are difficult to emulate, thus extraction of the relevant information consumed considerable time and effort.

6.1.4 Access to sources

Access to the data sources employed was provided freely upon request. This was usually because of existing relationships between the Geology Department and the body approached. Industrial sources would probably have proved more difficult. The promise of anonymity and access to the final data can often smooth the path to this particular door, especially now that there is a growing internal awareness of the waste of the data held by these companies.

Access to the computer stored information was not a problem because the same computer was used for both projects. If a different machine had held the older data and, as in this particular case, only compiled copies of the retrieval programs were available it is doubtful whether re-use of the data would have been attempted.

6.2 Statistics On The Initial Data

6.2.1 Results of distribution checks

The spatial statistics of the source data and the effect of the sample distributions on interpolation routines is discussed in Appendix 1. It has been proposed that the averaging of the information in the map onto a regular grid will improve both the efficiency of storage and the accuracy of the interpolation programs. Map segment, or pixel, sizes from the untreated ten metres, provided by the original coordinates, to 200m meshes were studied in the tests. The data used were derived from Varey's (1977) databank. For a well documented area of this information the fifty and one hundred metre grid meshes achieved 14 and 25% coverage respectively.

Two separate statistical tests showed that the 200m segment size reduced the source information to a uniform, random distribution. The 100m segment also offered a random distribution but this was not found to be uniform. The averaging on to a grid caused certain effects in the surfaces interpolated from these new data. Surfaces were often smoothed with respect to the original and there was some shifting, or translocation, of features because of the alteration in the spatial distribution of the data.

The 100m mesh performs well in areas of high data density, where it offers a slight smoothing and low translocation. It is also very good in areas distant from any data points. In areas of low, uniform data density there are some translocation problems. However, in these areas very little is known about the surface and the use of the smaller segment sizes would lead to a great waste of file space for very little real gain in accuracy. In addition the 25m and 50m meshes offer lower accuracy in areas of high data density, due to spatial averaging effects. The 200m mesh has an unacceptably low resolution for the features which are expected to occur in a urban environment underlain by glacial, postglacial and man-made sediments. Thus the choice of a 100m segment for the database can be justified not only on grounds of convenience but also by statistical and empirical tests.

6.2.2 Selection of parameters for use in the map

The classification of rocks and soils, as defined for engineering purposes in the geotechnical standards, runs as follows:

colour, grain size, texture and structure, discontinuities within the mass, weathered state, alteration state, minor lithological characteristics, ROCK NAME, estimate of strength, estimate of mass permeability, other terms for special characteristics.

The geotechnical test results are summarised into this definition.

Varey's borehole storage system catalogues all the information which may be taken in a normal survey and places it in ten separate files, containing over one hundred geological, geotechnical, chemical and hydrogeological parameters. Table 6.1 presents the occurrence of the most

Table 6.1 Statistics of Varey's Borehole Data

occurrence of GEOTECHNICAL parameters taken for 5052 records.		occurrence of GEOLOGICAL parameters taken for 2557* records.	
PARAMETER	%	PARAMETER	%
Standard Penetration	66·2	lithological type	99·8
moisture content	29·8	stratigraphy	99·6
bulk density	16·1	lithology	79·8
shear strength	15·9	genesis+	78·6
pH of water and soil	9·9	primary colour	65·4
sulphate of " & "	9·9	strength estimate	28·9
Atterberg limits	7·6	second colour	9·2
grading curve analysis	6·3	weathering	2·3
consolidation	1·5	texture	0·7
specific gravity	<0·1	fabric	0·5
CO_3 and Cl^-	0·0		
A score above 10% indicates that a parameter is likely to occur at least once in each borehole.		A score above 5% indicates that a parameter is likely to occur at least once in each borehole.	

* The total number of records is >5000.

+ This parameter is derived from the lithology and not specified in the original data.

commonly recorded results in his 522 coded boreholes.

Clearly if a simple fixed-format storage method is employed and if all the parameters are recorded there will be a large amount of wasted space in the datastore. A balance must be struck between exhaustivity and efficiency. The primary use for the proposed map database is to provide information, as a map, on what is there, rather than the transfer of any great depth of information. Therefore, it was decided to use all the parameters which stand an even chance of occurring at least once in each borehole.

The Atterberg limits fail this test but are so useful, especially in conjunction with the moisture content, that they have been included. Weathering also fails the test but was considered to be essential to the description of rocks, particularly the Keuper Marl, near the surface. An extra qualifier for the geology, the "Primary Mineral", has been added to provide more sophistication in the lithological descriptions. This parameter describes unusual minerals, organic inclusions and characteristic trace features such as gravel in a clay or the slight clayeyness of the Upper Mottled Sandstone.

6.3 Initial Inputs

The format and codes for the data recorded are listed in Appendix 2. Reasons for the decision to use only equivalent look-up rather than calculated percentage codes for parameters such as the colour and rock type are given in the preceding chapters. For efficient, error-free data capture the coding techniques must be simple to use and the data must be in a form which is easily inserted into the computer, as well as readily readable by reduction and verification programs.

6.3.1 Recording techniques

The only route onto the computer, for large amounts of data, which was available at the start of the project was the card reader. Punching of cards is carried out by an operator who has no knowledge of geology. Griffiths and Rosenfeld (1954) found that the number of errors made by punch operators is at a minimum whilst typing numeric codes. They surmised that this was due to the absence of verbal implications in the data. Another major source of corruption which can often be seen is the misplacing of lone digits to the left or right of their specified location on the card. This leads to incorrect input to the fixed format data reading programs employed to make the best use of the space available on the input cards. In an attempt to reduce these two sources of error the codes used are wholly numeric and the datafields on the input forms are compressed, well delineated, labelled and, as far as possible, of fixed widths.

Only two coding sheets are used, to minimise the problems of cross-indexing between sheets. The header sheet contains the level and test data and the second sheet is for the geological layer information. Both sheets have the number of the borehole in that particular segment square and the level of the base of each layer, for each layer. This provides three correlations to check in addition to the simple order of the records. On the header sheet there is room for one value for each engineering parameter, thus the number of cards depends upon the largest number of test values. The geological layers have one card each.

The single value per parameter per borehole method facilitates simple entry of the generally sparse engineering data but for boreholes with many results for several samples pre-averaging is allowed, to reduce the number of lines - and therefore cards, required. A special value is entered, to indicate to the card reading program that this pre-averaging has taken place. Detailed information on the codes and formats employed is given in Appendix 2.

Throughout the database all levels are taken as heights above Ordnance Datum. If a borehole is encountered which has no recorded O.D. level the relative levels are calculated at a later stage from the digitised version of the present day terrain held in the datastore.

Site investigation information was coded at the Structural Engineers' offices. Their reports are filed by record number and referenced on 25 inch to the mile plans of the city showing the location and extent of the sites. Coding is thus made easy. One scans the 100m square to be mapped on the plans and extracts the relevant files. The reports are then translated on to the coding sheets, in Metric or Imperial units, and the search continues. If new codes for encrypted parameters are required they can be added to the code lists as and when necessary. Because reports vary from two layer 1925 trial pits to 45 borehole road surveys, with full test results for each layer, it is difficult to determine a more specific coding time than an average of less than five minutes per borehole.

6.3.2 Computer inspection

Computer verification and conversion to S.I. units of the punched information took place after the full files were stored on the computer. A more modern method of data capture would be better suited to this type of information, to provide instant checks of the input and, possibly, recall of the other data available for the current segment. This could be achieved with a data logger such as a small, or even pocket, microcomputer and a cassette tape recorder. The pre-checked and reduced data could be transferred to a larger machine, via a standard interface, for bulk storage at a later date. A fairly simple program in the logger would suffice to make this new technique better than the use of coding sheets. A sophisticated routine which translated codes and converted units could improve the coding time by several minutes and markedly reduce the number of errors.

6.3.3 Varey's databank

Conversion of the information from the existing computer stored files to the same code standards as the new data proved to be a complex task. The first stage was a check of the reliability of the pointer structure of the databank, its layout is shown in Figure 6.2. The files which are required are marked thus *. Information is held in a fixed format. The location in the card image identifies the parameter.

READ statements, to extract the relevant information and mask the rest were constructed from Varey's descriptions. A preliminary scan program checked the correctness of the subordinate file structure, mimicking the operation of the translation in a faster program. Due either to inaccurate initial input or subsequent corruption a number of pointer errors were discovered. These included two 'extra' boreholes which had lost their header cards.

Translation of the colour and lithology to simpler

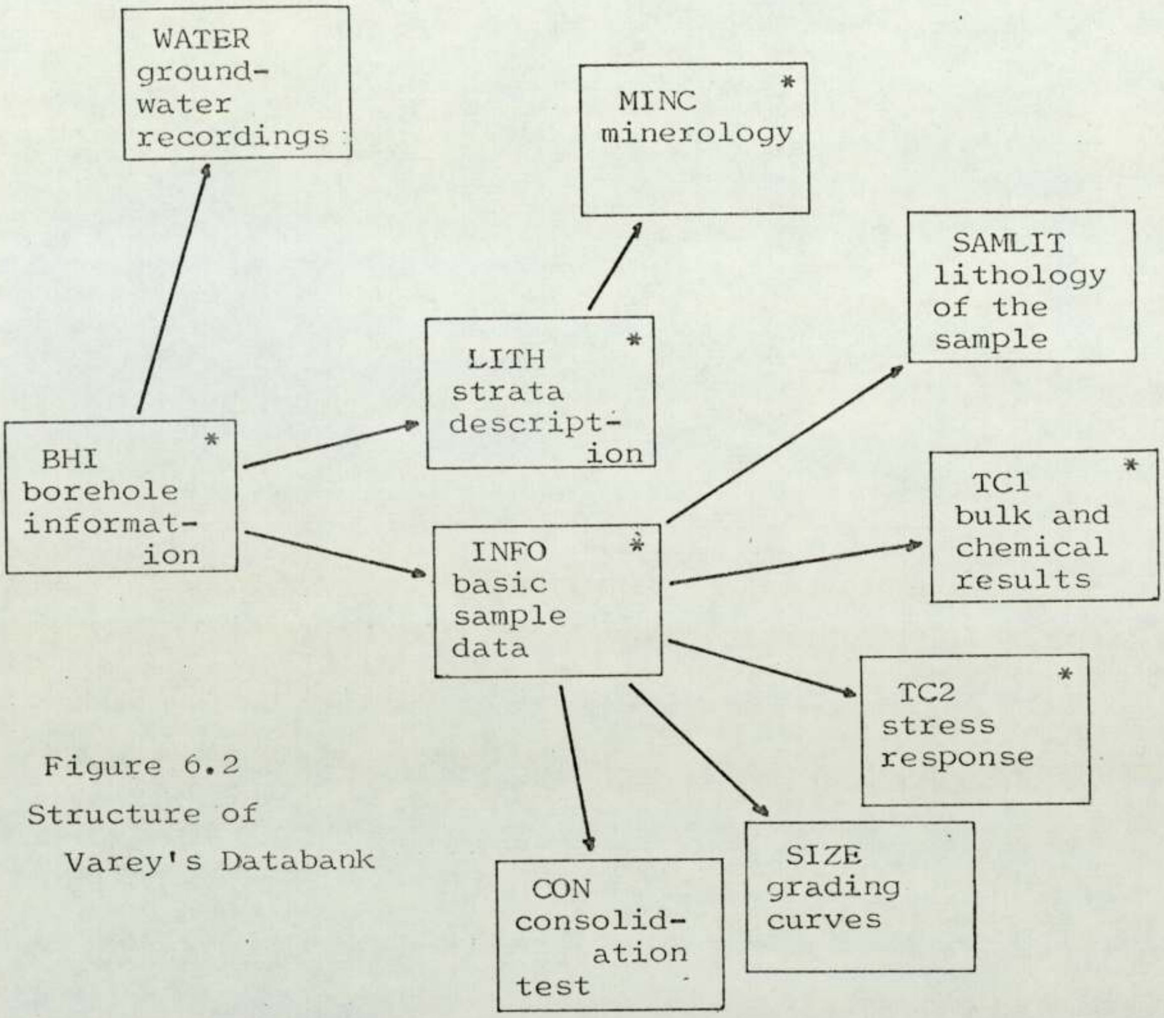


Figure 6.2
Structure of
Varey's Databank

codes were performed after correction of these errors. Colour codes were determined by a subjective calculation based on the HSL colour solid. This was an adequate solution since the initial coding was from verbal descriptions and did not contain the nuances of which the HSL technique is capable.

Lithology presented a more difficult problem. Varey's routines and the ICL GEORGE editor were used to make up a file containing the verbal descriptions of the lithologies. These were translated to numeric codes and the errors produced by Varey's program listed by a check program. These errors were 'corrected' using the GEORGE editor.

A final program read all the relevant files, converted the levels from depth in hole to heights O.D. and output the information, in the same format as the verification program for the newly-coded data.

6.3.4 Location implied data

The implied location data were collected by placing a grid over the required map on a 100m square mesh, scanning across the grid and transferring numeric codes for either the stratigraphy or the spot/contour height encountered to a gridded piece of paper. These lists of codes were then typed into the computer either line by line, for the stratigraphy and hazard data, or by the kilometre square box for the level map. The level data were interpolated to provide a full digital terrain model (DTM) for the present day topographic surface of Birmingham, using the Shepard (1968) weighting function.

The Water Board information was translated to codes and added to the general hazard/archive grid, its location specified by National Grid coordinates. This grid also contained information collected during the study of site investigation reports and archive maps.

A program, SPLIT, has been provided in the final map database which allows the conversion of a few parameters of coordinate located, such as archive, data to be inserted as a list, rather than as a full map. It should be noted that the entry of point location and vector data, such as road and isoline maps, would be made much easier if some form of digitiser was available.

6.4 Computer Aided Verification

6.4.1 Pointer checks

These were performed on the ICL 1904s during translation of the initial data. A very small, fast scanning program, running under high priority in the background stream, detected coordinate or character errors and pointer mismatches. The program halted after recording the line numbers reached (i.e. before the error was detected) and indicated the location and type of the problem. The relevant records were studied, by the operator, and a solution to the problem, either omission or correction, edited in. This was obviously a time-consuming business, but data checking is a ubiquitous (and frequently 'glossed-over') step in the entry of complex data to computer systems.

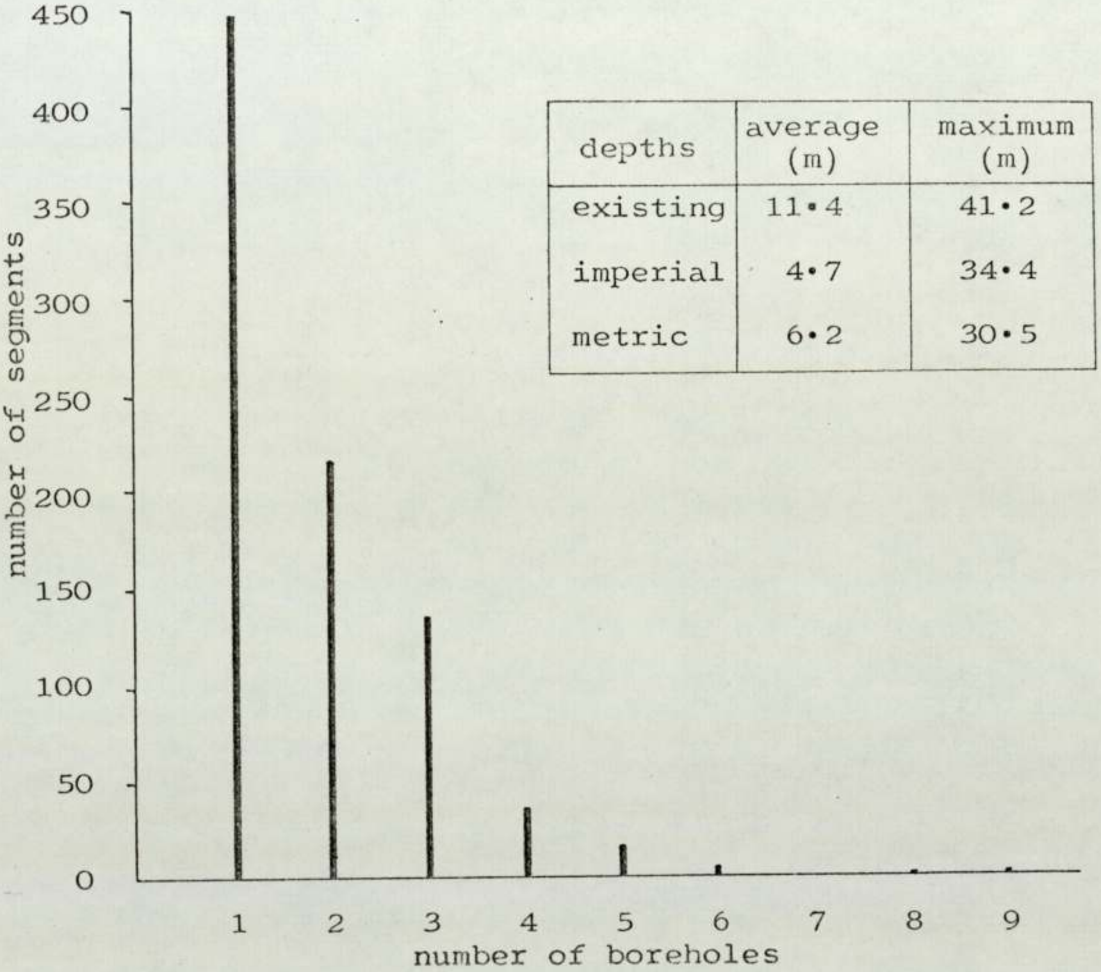
When the data could be scanned without errors, which would cause a crash, the three sets of: new metric, new imperial and converted existing files were written to three intermediate files. These contained all the necessary information for the final synthesis in a part-processed, standard form. The order of the boreholes, first by X and then by Y coordinate, across the three files is held on a 'log' file. The format of the intermediate files is given in Appendix 2.

6.4.2 Conversion, translation and statistics of the data

All of the prepared files were transferred to the newly-installed Harris 500 computer. The site investigation files were analysed for the scores of the different parameters coded. Any unlikely values, such as scores for unassigned codes, were noted and the most likely recording or typing error estimated. The scores for the real values are included in Appendix 2.

The total number of boreholes recorded is 1546. The maximum number of boreholes in one 100m segment is nine. The average number of lithological layers per borehole is 3•94 and the maximum 25. The distribution of boreholes per segment and information on the depths proved in the boreholes are given in Figure 6.3.

Figure 6.3 Numbers and Depths of Boreholes



depths	average (m)	maximum (m)
existing	11·4	41·2
imperial	4·7	34·4
metric	6·2	30·5

It is difficult to estimate the percentage of coding errors in the site investigation data. Those which appear in the tests are much less than ½% of the total number of recorded values. Because the map will eventually be made up from averaged values this low proportion of errors does not justify reference to the original data for correction. Instead, a subjective replacement of the obvious errors with an estimated most-likely value is used. The pointer errors encountered in the initial data were corrected and do not contribute to the total.

Combination of the site investigation files and the location implied data took place as a large background job on the Harris. The errors detected in the codes and conversion of the special recording techniques, such as the negative relative heights for boreholes without a specified O.D., are accomplished by this program.

6.5 Experiments With Two-Dimensional Pattern Recognition

6.5.1 Data quality

The synthesis of the integrated typical segment information from up to nine boreholes per segment in a study area of great lateral sedimentary variability is a complicated problem. In addition to the fluctuations in the soil type, there is the major problem of the quality of the site investigation information. This is often related to the old problem of 'experiments reveal that what the geologist perceives in and remembers of rocks is not the same as what is actually there', Chadwick (1975). These data have also accumulated over many years, using several measurement systems and often with little regard to standard formats. Each investigation is designed for a specific requirement and a specific site. The early investigations use local terms for soils and rocks, whilst many recent studies make no differentiation between the two. Borehole records and results are filtered through the limited language and variable interpretation of the 'standard' report. These effects lead to "noise" on the terminology employed for particular rock or soil types and colours. Stratigraphy is rarely specified for the softer components in a section.

6.5.2 Problems of totally computerised methods

Attempts to reduce such records by purely computer-based techniques lead to a number of experiments with the calculation of identifying combinations of parameters for each layer. The first experiment, called ALIGN, tried to re-level all the boreholes in a segment by averaging the levels of various lithological or stratigraphic divisions within the set of boreholes for a segment and was partially successful.It employed a sliding scale of typifying parameters from specific colour to broad stratigraphy. Correlation between the boreholes, however, was rarely achieved at a lithological, rather than a stratigraphical, level.

The next stage was to merge the separate boreholes into one archetype, whilst recording the worst case values. Several methods were tried out for this process. These all employed a scan line which moved down calculated mathematical sections of the boreholes 'deciding' whether the newly encountered set of values were sufficiently different from the current synthesised layer to justify the creation of a new layer. Some of the rules for layer discrimination were as follows:

A simple selection of the commonest lithology and colour with an associated stratigraphy.

The addition of a worst case weighting to the above parameters.

Weighting without averaging the parameters.

All of these discriminations lead to a very reduced number of layers on the synthesised section. This results in a loss of detail which does not do justice to the information available in the original boreholes.

The final attempts included a weighting for how close the scan was to the centre of a layer and how 'good' or 'bad' that layer was. These discriminations lead to too many layers in the final synthesis and an "ungeological" section.

There is no doubt that with a well controlled set of data on flat, fairly uniform strata, if sufficient time and effort were expended, this computerised merging would be possible. But, in the time available to the project and with the data collected in a numeric code format - despite its generic-specific nature, another method had to be found.

6.6 Computer-Aided Processing

6.6.1 Theory

The inappropriateness of totally computerised techniques for the reduction of variable quality and quantity data to typical sections for each 100m segment suggests that a more flexible approach is required. The human brain can instantly resolve very complex spatial relationships, and geological data from borehole records are of a very visual nature (Bouma & Nota, 1961). A manual synthesis process, using the computer's ability to display information, has been developed. Figures 6.4 and 6.5 show this process in a schematic form and the following sections describe it in some detail.

6.6.2 Borehole synthesis

A display program, MERGE, which works with the colour graphics terminal was designed to show all the boreholes for a segment. These were drawn to scale, side by side, in full colour and synthetic texture with an indication of the stratigraphy. Other lithological parameters were available on request. The operator was required to select 'similar' layers in the set of boreholes. The computer automatically processed the data for these layers into a new synthetic layer. The best represented values for parameters were assigned to the current synthetic layer and the level of the base of the layer was calculated as the average of the

Figure 6.4 Computer programs in the synthesis

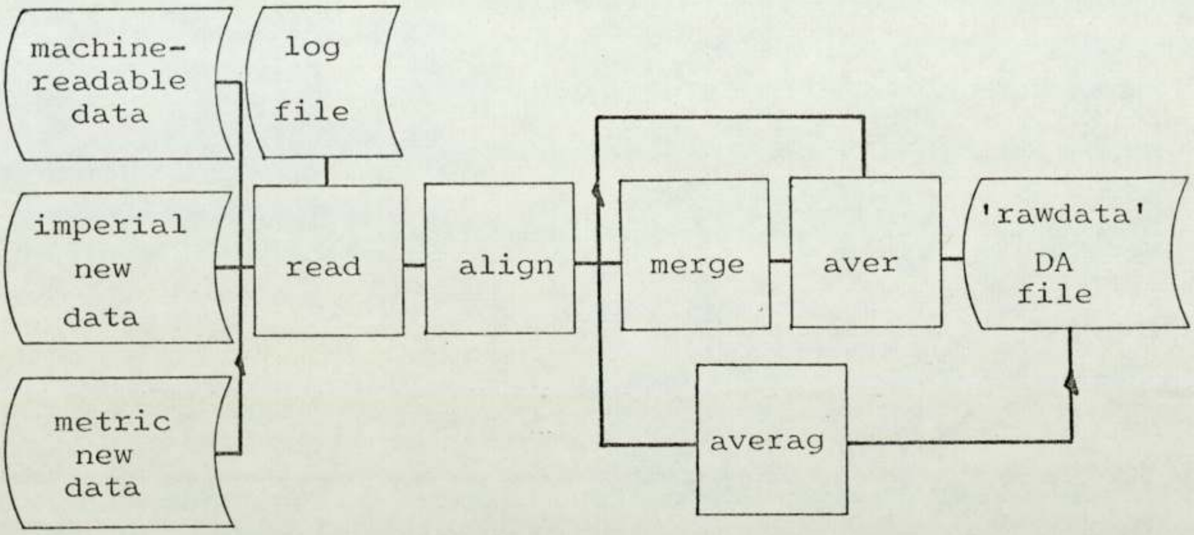
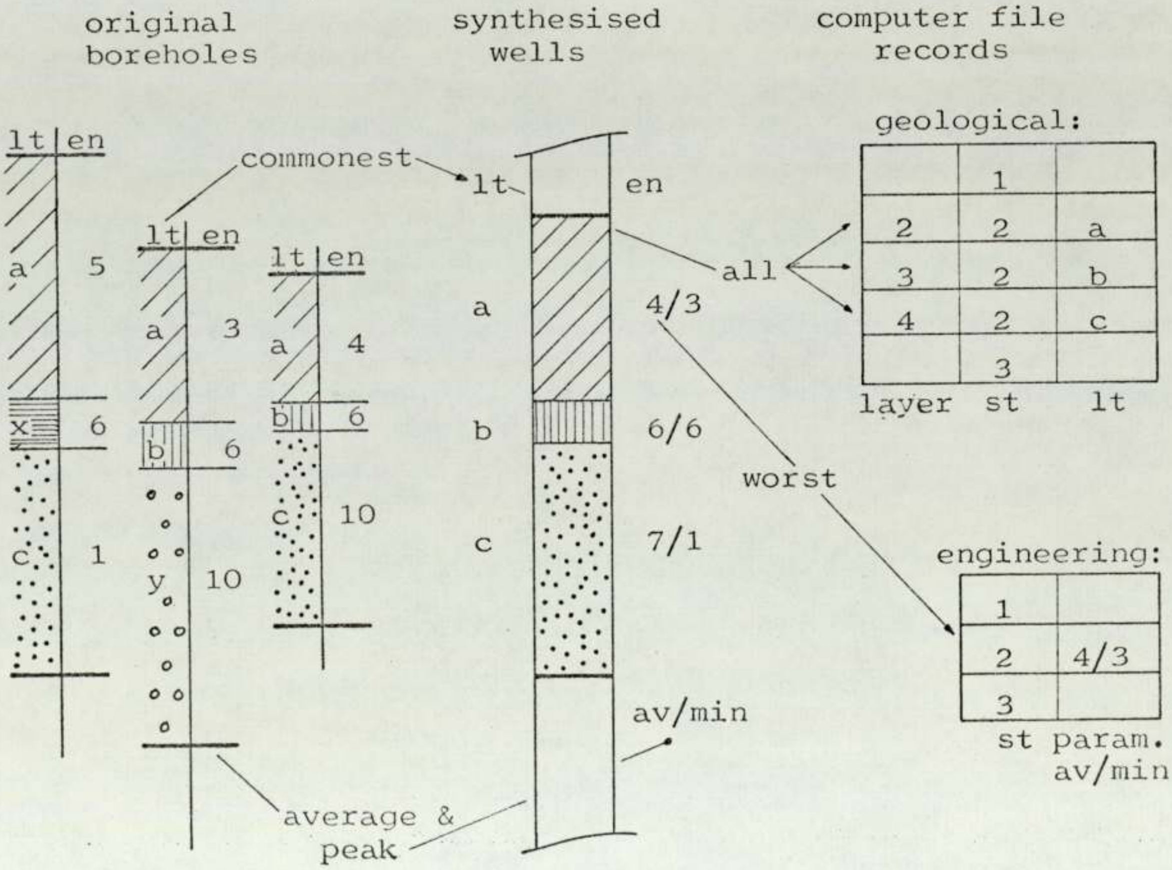


Figure 6.5 Schematic summary of the synthesis



st — stratigraphy lt — lithology av — average
min — minimum param. — parameter en — engineering

levels of the bases of the selected original layers. Obviously a suite of housekeeping and manual overrides had to be included in the program. The output from the process was written to a new serial file with a cross reference table for the averaging of the engineering results to the correct layers of the synthetic borehole.

The philosophy behind this technique is the fact that a best colour, texture, location guess is often very near to the best possible spatial analysis solution and, as long as the display is not concealing or misrepresenting the information, a reasonably good approximation to an archetype for each segment can be achieved by an 'instant' impression.

The resultant segments' lithological records of archetypal boreholes are referred to as 'wells'. This term is intended to differentiate between original boreholes, the full map segments, which include archive and extra level information, and the vertical section information of a segment.

6.6.3 Checking and the addition of the engineering data

After the synthesis another program, AVER, allowed the user to check the final result against the original boreholes. This program provided a range of correction and fine tuning tools, such as the ability to adjust the level of a layer boundary or enhance a low strength result. The program wrote its output directly to the direct access datastore files. It also averaged and processed the engineering data. Later, however, a more detailed division of this information was added via another program, AVERAG.

The engineering results for each synthetic layer were calculated from the average of all the values which were recorded for the original layers and were assigned to that synthetic layer. The most extreme of these original values is also preserved, as are the total number of values and the layer number. The final stage in the process examines the sets of results for each parameter for each layer in stratigraphic groups. The worst case value of each parameter's average value, for each stratigraphy represented in the well, is used to make up a composite 'worst case, average' set of engineering results. The number of values, extreme value and the synthetic layer number associated with the selected average value are also stored in the final engineering data record for that stratigraphic level.

6.6.4 Correction of the datastore

The total number of segments containing site investigation information is 854. This provides an 18% data coverage of the mapped area. This information is held in the final datastore which is amenable to checking by the map enquiry program EXTRACT. Checks were made for unassigned stratigraphies, which were left blank specifically for this stage of the operation, and for anomalies between the ground level of synthesised wells and the O.S. topographic DTM. The reassignment of doubtful stratigraphies, via the EDITOR, could now be made by examination of the adjacent segments and the exercise of prior geological knowledge, or ground truth. Additional tools, such as the recorded depth of the bedrock against an interpolated surface value, could also be employed, for both the re-levelling and the stratigraphic corrections.

Re-levelling was only carried out on a few segments, and then only with caution and due regard to the: number of original boreholes, the existence of made ground and the presence of excavations. It was probably made necessary because of poor surveying or mistyping of records or by a large topographic variability in the segment concerned. Wherever possible the wells' levels were tied to the more reliable, but less accurate, bedrock surface.

The late stage checking and correction, whilst being highly subjective, was controlled by an objective 'search for errors and stop' program which filtered out the system 'noise' and allowed only gross errors to be brought to the attention of the user. This both speeded up the checking procedure and prevented the possibility of the 'intelligent user' redesigning the whole map according to his personal preferences.

7 A USER'S MANUAL FOR THE MAP DATABASE

This is a description of the functions and uses of the datastore and its manipulation programs. A user's model for the 'map' is introduced and the various operations which can be performed on it, in its implementation on the Harris 500 computer, are explained.

7.1 The System

Figure 7.1 shows a conceptual model for the computer-based map. It illustrates the three forms of information output from the system namely: single segments, sections and maps. These may be read and examined on either a Visual Display Unit (VDU) or paper copies.

Information is held in the database as 'segments'. All the available data for each 100m square of the datastore area is recorded in a segment. Each segment is known and addressed by its National Grid (NG.) coordinate. The SECTION program displays selected information from a row of segments and maps are selected individual parameters taken from all the segments in the area.

Access to the map database is via the Harris 500 computer and a Sigma T5680 graphics output controller situated at Aston University. A description of the Harris and its VULCAN/VOS operating system and job control language (or JCL) can be obtained from the general introduction and JCL manuals issued by the Computer Centre. If the system is not online a request must be made to the Computer Unit or the computer representative in the Geology Department at the University.

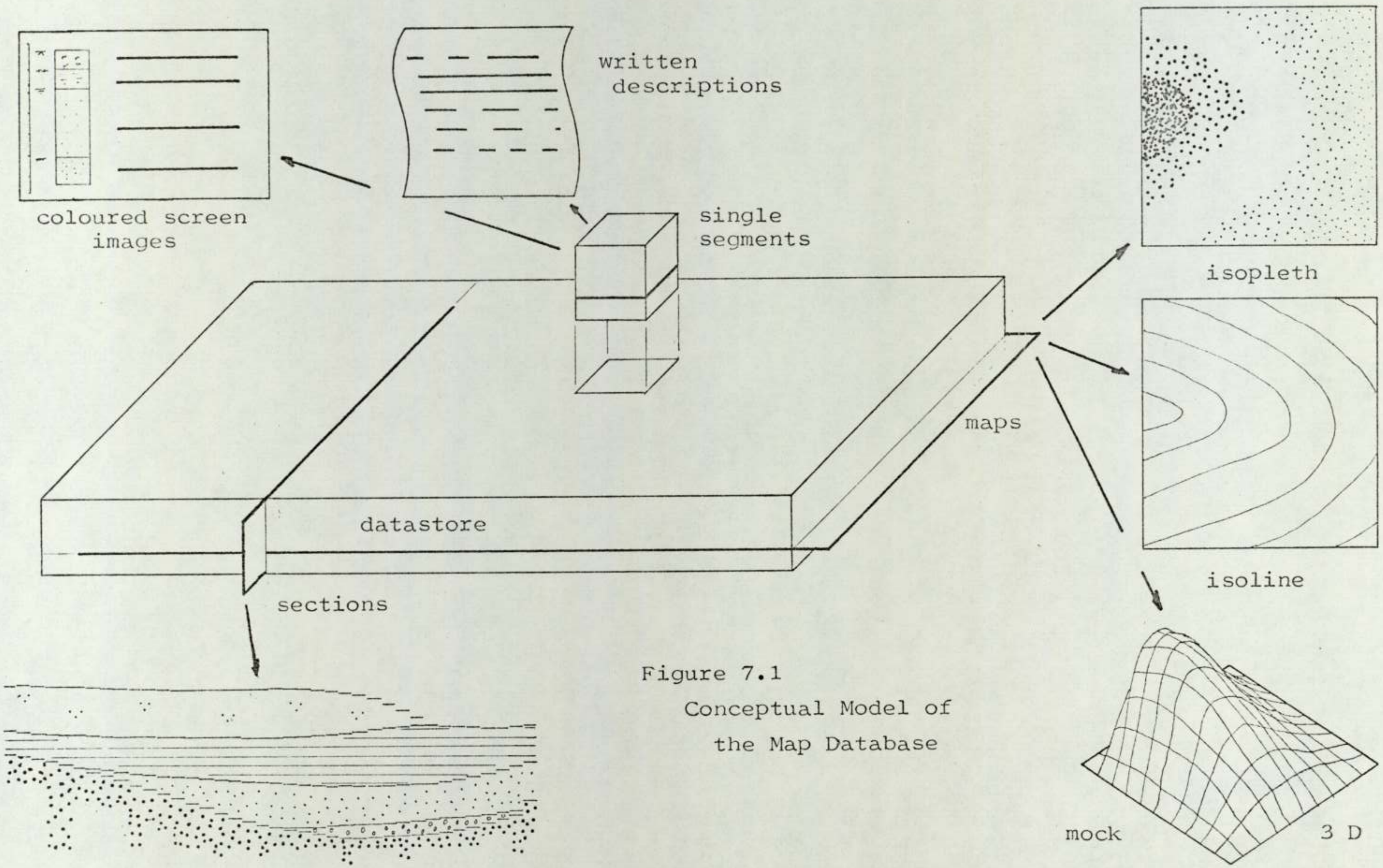


Figure 7.1
Conceptual Model of
the Map Database

7.2 Session Ordering

After the user has logged-on to the Harris, on either an ordinary Newbury terminal or a storage tube simulating terminal (such as a Newbury 8005 or DataType DT22) or the Sigma T5680 colour graphics terminal, the command STARTUP should be issued. This will activate the main system programs. All the programs are designed to operate upon the typing of their names prefixed by 'R', e.g. RREPLACE for REPLACE or RSCATTER for SCATTER.

If a number of maps are required and the workload on the computer is high it will probably prove necessary to have two sessions on the terminal. The first of these is devoted to issuing the necessary retrieval program to extract the map data and the second, probably a day or so later, to use the information selected by this slow background program. Output to the wide-bed CalComp drum plotter will only be plotted at night and will therefore not be available until the day after it is requested.

7.3 File Types

Apart from the computer data storage files holding the programs and macros* there are five types of file used by the map database.

7.3.1 Information

Three files of text:

MAPINFO This describes the use of the Harris and the database programs.

---

* Programs are machine instructions written in a language such as FORTRAN. Macros are programs of instructions for the operating system, written in the JCL of that system.

CODEINFO The numeric equivalent codes employed in the datastore and their text equivalents.

PROGINFO General programming standards to aid elucidation of the program listings.

These files are standard sequential files and may be viewed using the standard JCL commands LIst and DIsplay.

7.3.2 The datastore

This is described in Chapter 4. It consists of four files:

FULMAP An index to the store which gives a spatial sense to the information.

FULHEAD The master file containing the pointers, levels and other general information.

FULLAY Detail for the geological data.

FULENG Detail for the engineering data.

These files are of direct access type and can only be read by special programs. A suite of suitable routines is provided in the service programs for the database. When the REPLACE program is used a set of new datastore files must be created. The prefix to their name is not FUL-- but NEW--.

7.3.3 Rundata

A loose family of files used to transfer electives to the batch programs or to register values required by different routines, e.g. the video lookup tables.

7.3.4 Maps

Fixed format serial files which may be used to save extracted sets of spatially related parameters outside the main datastore, for faster access. The maps are a standardised medium for data transfer between the processing programs.

7.3.5 Workfiles

These are generally maps, but are not permanent in type. The operating system has a number of special files which are held outside the filestore space allocated to the user. These are erased when the user logs off and the information in them is therefore lost. Because it is convenient to work with this type of file the map database system uses workfiles for output. However, if the files are required to be preserved between sessions they must be renamed, with names other than those of workfiles. The serial workfiles used by the map and operating system are named W1 to W9 whilst files B1 to B9 are generally available to users.

7.4 Segment Processing

Single segment records whether original, synthesised from a number of boreholes or interpolated from adjacent map segments may be viewed in two ways: EDITOR shows a schematic representation in colour and synthetic texture whilst SEGMENT gives a numeric and verbal description. EDITOR allows alteration of the information in a limited fashion, to respecify all of the parameters, to correct errors or enhance the information. Complete new segments may be inserted via SEGMENT, if they are first converted to the codes employed in the datastore.

7.5 Section Display

Geological sections through the mapped area in colour, synthetic texture and an indication of stratigraphy are displayed by SECTION. Additional information, such as the water table, can be added to the diagram at will. Scaling, to make the best use of the drawing area, is automatic.

7.6 Map Processing

A large number of different sets of information may be withdrawn from the datastore as maps of the whole area by EXTRACT. These maps are in the form of files which can be displayed and processed by a number of routines.

Processing of the maps may be either mathematical or logical. COMBINE provides for simple one segment at-a-time operations, i.e. simple image processing. COMBINE has a number of standard functions and a facility for the user to insert new processes. Intersegment operations (image analysis) are limited to interpolation of data for unfilled segments in the maps by either EXTPOL, for real numbers, or PXTPOL, for integers.

The data distribution for the maps can be displayed by SCATTER. This program also allows the range of values in the map to be freely divided up into classes by the user. This division facilitates the display of maps in an isopleth form, thereby aiding detailed spatial analysis.

Map data can be displayed as a fast VDU or line printer isopleth map by CHORO, which also allows the image to be bit sliced. Alternatively, the map may be shown in high definition as isopleth or isoline plots or a three dimensional projection via DRAW. The latter program allows several maps to be superimposed, as an aid to elucidation of the total structure of the mapped area.

7.7 Updating

There are three reasons for an update to the datastore:

INEFFICIENCY due to a large number of insertions which change the datastore structure.

RESTORING of the direct access files from a serial file backup listing after corruption or transfer to a new location.

ADDITION of a new parameter or file to the datastore.

An update creates a new set of files but the old files should only be erased after checks have been performed to ensure the correctness of the new files. The best checks are probably extraction of a complex map and the listing of segments from the bottom left and upper right hand corners of the map.

7.8 Notes Of General Guidance

7.8.1 Use of the break key

The programs in the system are not designed to be fail-safe, except in protecting the datastore and themselves. It is therefore recommended that the philosophy employed in learning about the system, and approaching the processing or enquiry, is to rely on the 'break' key to stop and then re-issue the program, rather than to try to recover from errors. The background programs will not run and graph plots will not be issued unless an orderly exit is made from the calling programs.

7.8.2 Graphics and the graphics cursor

The graphics programs on the T5680 terminal use a CURSOR. This is a white cross displayed on the screen. It moves in response to the joystick or inching buttons and a 'hit' is achieved by pressing the button marked HIT in the centre of the inching buttons. Because of the mode of operation of the graphics terminal <u>no other key</u> must be touched during interactive work with the cursor unless it is prompted for by the computer. All the programs except the illustrations, SECTION and the EDITOR may be used on standard or DVST terminals, as well as the Sigma terminal.

7.8.3 System standards

Two standards must be noted:

The stratigraphy codes are divided so that the first number in every ten is a generic descriptor which the system understands as a reference to that whole stratigraphic group. i.e. Generic: 0 - Recent, 10 - Glacial, 20 - Man Made, 40 - Permo-Triassic; and the more specific: 42 - Keuper Marl, 14 - Hoxnian.

The terms: prox(imity) zone, array or map and reliability map, index or level all refer to numbers which, if positive, indicate how many values are averaged into the current value and, if negative, how far away the nearest real information is to that interpolated value. The proximity classification is as follows:

prox. value	-1		-2		-3		-4		0
range (m)		100		200		400		800	

N.B. the zero and positive values give special meanings in the chemical and penetration test results.

7.8.4 Levels of expertise

Three levels of use of the system may be defined:

BEGINNER. Employment of the MAIN programs, usually only for one-shot hard copy. Simple maps with limited use of the standard COMBINE and SCATTER functions.

INTERMEDIATE. Main and SERVICE programs for sets of maps, using the custom ECOMBINE facility and PSOUT. Insertion and editing of new data, via SEGMENT and EDITOR. Simple use of REPLACE and the interpolators.

ADVANCED. Use of custom sections in EXTRACT and REPLACE for extensions to the database, generated via SPLIT. Use of DRAW at different scales on new devices, such as the microfiche plotter at Manchester. Reinterpolation of the data with better or different laws. Creation of new search indices.

In addition to these levels the recommended enhancements, in Appendix 3 and the conclusions, could be implemented by a third or 'fourth' level user.

7.9 The Main Programs

7.9.1 Interactive programs

7.9.1.1 COMBINE to run RCOMBINE output to W1,8,7

The routine prompts for data maps and the associated proximity arrays. If proximity maps are not required any map name may be supplied, providing the file is not a data input to the current program. Up to four maps may be attached. The attachment process may be terminated by typing END at any prompt.(This is a standard technique for the attachment of map sets in the map database system)

When the program starts it prompts for the function required from amongst the following:

1 addition

2 multiplication

3 percentage difference

4 consistency index

5 a map of the largest value from two input arrays

6 a decoder for maps containing multiple parameters, such as the archive data

7 map of the contents of one map depending on a certain value occurring in a second map

8 alteration of values greater than or equal to that specified

9 alteration of values less than or equal to that specified

10 conversion of a proximity file for use in a display program

11 a custom section which can be inserted before running

The program will prompt for electives and extra information, such as the scale factor chosen by the user for the output map to keep the output data within the four significant figure map resolution.

Custom sections may be used by placing the required function, written in FORTRAN 66, in the file BIT and entering the command ECOMBINE. This will prompt for the length of the file and edit the section into COMBINE. The four input files should be represented by the variables A,B, C,D, the two constants as C1 and C2 and the outputs as X and Y. For convenience the input and output maps are all defined as real.

7.9.1.2 SCATTER to run RSCATTER output to W2,9

The routine prompts for the map to be analysed and then whether zero values in the map are to be scored as actual values. A scattergram, or distribution histogram, is plotted on the screen. The user is invited to define class boundaries by typing '1' characters below the desired values. These '1's should be separated by either spaces or '0' characters. Merely typing 'return' will bring about the next prompt for four electives to:

end the run of the program

plot a copy of the histogram in file W9 for listing

integerise, or classify, the data held in the input map, according to the pointers given, and write the output to W2

restart the classification

The program, when it ends, will ask if the scattergram, if produced, is to be plotted. This paper copy will, in addition to the screen information, have a horizontal scale and vertical scoring information.

7.9.1.3 CHORO to run RCHORO output to W3

The routine prompts for the colour table (as described in THREE and COLTAB) to be used and the map file to be plotted. It displays an integer or a classified map in up to 15 colours as a choropleth of regular squares with an annotation of the class intervals. A prompt is made for:

the end of the run
listing the map to the line printer
bit density slicing

A label is requested and either bit-slicing electives appear or user-selected symbols for the printed map are requested.After the program has terminated there is a prompt for the printing of the map, if one has been produced.

7.9.1.4 DRAW to run RDRAW output to W6, PSEUDO

The program is compiled at runtime to a workfile, which is also used by the EDITOR. Because of this conflict, either, if the program has not been used or if the EDITOR has been compiled after the DRAW program (in the same session) the first prompt, asking if the program has been used, must be answered 'NO'. The program then asks for the names of first the real, then the integer files to be displayed and/or plotted.

After a request for the number of real files specified the program prompts for the output device to be used. This option may be changed later to any other device but the pseudocode generator. This is a special imitation device, which stores images (at a map scale of 1:25,000) in a file that may be plotted at a later date. The other devices available are:

the storage tube, or an emulator
the colour graphics terminal

the CalComp plotter, for which the picture is scaled at 1:10,000

the Tektronix flatbed plotter

This last route allows the preliminary use of the DVST emulator and a variable scale factor.

The next elective is for the type of plot and the selection of default or user-specified controlling parameters. A prompt for the length and contents of the title follows. If default parameters have been requested the plot is then drawn on the selected device. If the user has elected to alter the characteristics of the plot, a prompt for the colour of the diagram is followed by a description of the various parameters of the current plot type which may be changed, plus an example of the necessary input.

A detailed description of the options for the Ginosurf routines may be found in the relevant manual. The isopleth, or symbol, maps prompt for:

the type of labelling and border for the map

the number of integer maps connected, to be plotted simultaneously

the numeric equivalents of the colours to be used for successive maps

the type of shading to be used for the each map

If character symbols are to be redefined then the new symbols should be entered on a new line and separated by spaces. After the plot has been completed there is a choice of:

a new plot type

selection of a new output device, or termination

loading of a new map of real data

clearing of the drawing area of the current output device

The last option is essential for the DVST because without it the new plot is superimposed over the previous one and while other devices have the necessary discrimination the storage tube display becomes unreadable. When the program has finished the user will be asked if he requires the plots sent to the CalComp to be plotted out.

7.9.1.5 SEGMENT to run RSEGMENT output to W5

The program prompts for the mode of operation required:

numeric codes
translation to standard English terms
insertion of new records

Numeric codes are implicit in the insert option. Once set into a mode a prompt is made for the N.G. coordinates of the segment to be listed. The information stored for that segment is displayed on the VDU.

If the program is in the data insertion mode, prompts for the new information will follow each line of the data listed. After all of the data has been entered an elective to place the new data in the filestore is offered. The location of the old records is sent to a file, OLDADDESS.

In all modes the program offers to print the list of the information for the segment to the line printer. It then asks the user if he wishes to change the mode of operation of the program or to end the run. Following this selection either a new set of coordinates is requested or the program finishes and allows the user to list the copies sent to the line printer.

7.9.1.6 EDITOR to run REDITOR (ref. DRAW)

This routine is provided as a service for the database but it will also provide a schematic display service for single segments. It must be used with the Sigma terminal. In display use the program simply requests the coordinate of the segment to be viewed and plots it. All other operations are driven from the menu. To get an invitation to enter another coordinate place the cursor outside the menu block, on the right hand side of the screen, and press the HIT button.

If values in the displayed segment are to be altered the operation required is selected from the menu by hitting the appropriate element. The program responds with the current value and an invitation to enter a replacement. To actually alter the datastore the UPDATE elective must be hit. Termination of the program may be achieved with or without updating of the currently displayed segment.

7.9.1.7 SECTION to run RSECTION output to W4

This routine is only usable on the Sigma terminal. An initial enquiry is made as to whether the user wishes to have a map array shown with the section. If a map is required the program prompts for its name and, after a few seconds, for confirmation of the connection of the map file. A prompt for a first cursor selection of an operation is displayed. The operations possible are as follows:

selection of a new section, on a standard vertical scale

regeneration of the current section with an optimised vertical scale plus the map and added level information

disposal of a copy of the section to the line printer

termination of the operation of the program

alteration of the background intensity

The 'new section' elective asks for several parameters:

the National Grid coordinate of the western or southern end of the section

the orientation of the section

the length of the section, in segments

The line printer option prompts for:

the width, in characters, of each column of the section
selection of compressed or open format
selection of the archive parameter to be displayed

At the termination of the program the usual request for the printing of the information sent to the line printer is offered.

7.9.2 Batch programs

7.9.2.1 EXTRACT to run REXTRACT output to named

When the command to use this (or any other background program) is issued a conversational elective preparation program is compiled and started. The data acquired from this routine are employed to control the operation of the real process, which is initiated at the termination of the conversational program.

EXTRACT prompts for a selection between two options:

0 extraction of only real information
1 all the information present in the datastore

The user is then shown a basic set of possible extraction modes which offer entry to the following sets of data:

1 parameters which have only a single value per segment
2 engineering information selected by stratigraphy
3 geological information selected by stratigraphy
4 geological information selected by height O.D.
5 engineering information selected by height O.D.
6 geological data from any layer in a segment
7 a dump to a serial file for study or transfer
8 jump to a custom section

Each of these modes has a secondary menu of possible types of information.

HEADER DATA

option one:

1 packed archive information and IGS maps

2 topographic DTM for the present day surface

3 levels of the tops of wells and overall reliability

4 worst case minimum height of the bedrock

5 worst rest water level

6 worst case minimum height of the made ground base and the worst type of deposit in the segment

7 most extreme pH (if prox. is positive the value is the sample's litho. layer, if 100 - value is for water)

8 highest sulphate concentration (prox. as pH)

9 lowest Standard Penetration Test result (prox. as pH)

10 worst case maximum shear angle (prox. as pH)

11 packed maps containing the hazard map and strat. maps

STRATIGRAPHIC DATA, the data are selected from the specific or generic stratigraphy specified by the user.

option two, engineering:

1 average and maximum moisture content

2 average and minimum bulk density

3 average and minimum shear strength

4 Atterberg limits

5 the worst case, average values for all the above for the whole segment

option three, geological:

1 commonest lithology

2 commonest primary colour

3 commonest qualitative strength

4 level of the base of the specified stratigraphy

5 level of the top of the specified stratigraphy

6 stratigraphy occurring below that specified

7 access to a custom option

8 specific stratigraphy for the specified generic

LEVEL DATA, the maps are taken from the level specified by the user.

option four:

1 lithology

2 primary colour

3 qualitative strength

4 stratigraphy

option five:

1 average and maximum moisture content

2 average and minimum bulk density

3 average and minimum shear strength

4 Atterberg limits

SEARCH, specialist functions mainly involved with mapping

option six:

1 particular lithology, if it is within a specified range of codes - entered run together as one four-digit value

2 stratigraphy treated as above

3 primary colour treated as above

4 qualitative strength treated as above

5 score from 1 to 4 for the worst soil type present

6 score from 1 to 20 for the worst strength

7 level reached within a specific strat., if the well ends inside that strat.

Almost all of the parameters are produced with a reliability map which gives a measure of the proximity of real data to that shown. A return to the first menu without committing the user to any selection is available from all the secondary menus. Each elective asks for names for the output files to which the created maps are to be sent. It then allows the user to either end the program or start another search.

7.9.2.2 EXTPOL and PXTPOL to run REXTPOL and RPXTPOL

These programs are run in the same way as EXTRACT. The preparation program prompts for the names of the data and data distribution files. It should be noted that data distribution files are simply proximity files without any negative 'interpolated' figures, such as are produced from EXTRACT under the 'real data only' option. After this prompt the program asks if a barrier map is required and, if it is, requests its name. The program goes on to request the proximity boundaries, if required, by asking for first the number of levels of reliability (the maximum is six) then the radii of the zones (maximum one kilometre) in descending order of size, in metres. A prompt is made for the number of points to be included in the average for each new point, the interpolation factor. The higher this is, the smoother the surface becomes. For integer maps (from PXTPOL) this value is set to unity by the program, thus any value may be given. Finally, the program prompts for the names of the output map and the proximity/reliability map. It then allows the user to either terminate the running of the program or start a new map interpolation sequence.

7.9.2.3 REPLACE to run RREPLACE output to datastore

The first stage in an update using REPLACE is to GEnerate the files NEWMAP, NEWHEAD U, NEWLAY U, NEWENG U. The 'U' signifies to the operating system that the file is to be of direct access type. These files will house the newly created datastore. Issuing of the RREPLACE command will then offer the user a choice between reconstruction of the files from a serial dump file, named SERIAL, and streamlining the running of the system after a number of SEGMENT insertions. REPLACE may be altered to add either a new field or a new file to the datastore. The interpolated data may be stripped off the datastore by EXTRACTING only the real data and then REPLACING the resultant file.

7.10 The Subsidiary or Service Programs

7.10.1 Direct access handling programs

7.10.1.1 READ4

This program reconstructs the spatial index map, FULMAP, from the header map. The program could be altered to create a new index for some other parameter. It must be compiled and consolidated (or 'vulcanised') by the user.

7.10.1.2 READ5 to run RREAD5 output to named

The program prompts for the name of the direct access file to be listed and the serial file to which the output is to be sent. It then asks for the length of the records in the direct access file (ref. the Harris FORTRAN manual) and the length, in records, of the file. For the map's datastore files this value may be obtained by LIsting the tail end of the file FULMAP.

7.10.1.3 READER to run DREAD

The program prompts for the name of the file to be seen. It then asks for the length and number of records in the file. A loop is entered whereby the user supplies a record number and the line is listed. If the line number entered is -1 the program terminates. The SEGMENT program produces the record numbers of the three files at the start of each VDU segment listing in the order: header, geological, engineering.

7.10.1.4 EDITOR to run REDITOR

This program has already been described under Main Programs. It may only be used on the Sigma terminal.

Electives are provided to alter all the records in a segment. These can be seen and experimented with by using the EDITOR program. Alteration of the database will only take place if an UPDATE or END UPDATE option is selected.

7.10.2 Illustrations of colour and texture

These must be run on the Sigma graphics terminal.

7.10.2.1 THREE to run JS RTHREE output to a VLT

The routine prompts for the name of the file which will receive the prepared Video Lookup Table (VLT). A VLT is a set of 16 colour codes. This VLT file should have been GEnerated beforehand. A display of the RGB colour solid, a spot showing a colour and a menu is given. The menu offers:

increase or decrease of red, green or blue intensity (lighter and darker boxes of the relevant colours)

redraw the solid's box

redraw the labels

end the program

send the current colour to the VLT

An invitation to enter the first R,G,B values as integers is printed. The specified colour is shown, in its correct place in the colour solid, as a dot. This dot may be moved, using the menu. It changes its colour to suit its position. Each hit increments or decrements the relevant colour by one unit.

The absolute values of the three primary colours are displayed after the first (background) colour has been sent to the VLT file. The colour spot takes this last colour, to allow comparison with the new value. After 16 colours have been specified the program ends automatically.

7.10.2.2 COLTAB to run JS RCOLTAB

The routine prompts for the name of a VLT and displays it to the screen as blocks of colour and a list of numeric values for the red, green and blue intensities.

7.10.2.3 COL to run JS RCOL

The program displays either the Sigma special or line printer equivalent, WPR'72, characters used for the lithology textures as a table with their numeric codes. The sets may be changed, or the program terminated, by following a menu.

7.10.2.4 MUNSELL to run JS RMUNSELL

The program allows the user to try out the subjective colour naming system employed in the database. A prompt is made for the numeric codes from the three columns of the naming list. Copies of this list are given in Appendix 2 and CODEINFO. The colour is calculated and displayed on the screen. Switching off is accomplished by entering a negative primary colour.

7.10.2.5 SHADES to run JS RSHADES

This program may be used on any DVST emulator and the small Tektronix flatbed plotter. It shows the total default symbol set which is used in the DRAW program. The set may be scaled to any size, to allow the user to test the appearance of the symbols on the device he is to use.

7.10.3 General facilities

7.10.3.1 STARTUP to run STARTUP

This macro compiles and saves the main programs of the system and lists a summary of the functions provided by the system.

7.10.3.2 SPLIT

This program must be compiled, etc. by the user. It allows up to three new map files to be created out of a list of coordinated information. Each line of the input must be in the form: National Grid coordinate accurate to 100m, parameter one, parameter two, parameter three.

7.10.3.3 PSOUT to run JS PSOUT

This program simply reads the pseudocode file PSEUDO from DRAW to a DVST or the Tektronix plotter. X and Y shifts, if required, are prompted for (the values are usually 30.0 and 40.0mm) and the stored diagram plotted out.

7.10.3.4 SCREENS to run SCREENS

This macro recreates the background screen images for the SECTION and EDITOR programs. These are normally held in files to save the time involved in re-calculating them at each run. If the files are erased or corrupted then this program should be used. The macro also compiles READER for the command DREAD.

7.10.3.5 LAMB to run JS RLAMB output to W4

The routine prompts for the name of the map file holding the surface to be hill shaded. The position of the illumination and the contrast of the image are requested. This program should be used iteratively in association with the CHORO program to achieve the best image for the current dataset.

7.11 Bugs

The shear strength is stored and displayed throughout, despite occasional references to the compressive strength.

The very first line of the verbal description in a run of SEGMENT has the words joined together.

There is an occasional fold-over of the tail end of the deepest segment in a rescaled screen image from SECTION.

The texture does not cover the whole area of each segment for certain widths of SECTION screen images.

The main register array in SPLIT should be named AR in the definition statement rather than IAR.

The pseudocode backend will not function if it is CALLed after another backend in DRAW, this is a GINO-F fault.

7.12 References

Aston Computer Centre, Harris FORTRAN 66 reference manual, 1980

Aston Computer Centre, Vulcan Job Control manual, 1980

Aston Computer Centre, a general introduction to the computing facilities, annual

Aston Computer Centre, use of the Sigma and other graphics terminals, <u>in press</u>

Computer Aided Design Centre, GINO-F user manual, 1976

Computer Aided Design Centre, GINOSURF user manual, 1981

Chapter 3, Structure and Philosophy of the database, <u>ibid</u>

8 EXAMPLES OF DATABASE OPERATIONS

(and an explanation of the map set)

In this chapter the potential of the map database system is illustrated by two examples of its use. A number of ancillary demonstrations and hints on the preparation of maps from the system are also provided. The technique of potential surface analysis, its problems and operation on a computer are also examined.

8.1 Interpolation Of Maps

8.1.1 Linear data

The choice of an interpolation function for the EXTPOL program is described in Appendix 1. This function is a mathematical expression governing the weighting given to the effect of the value of a local data point on the point currently being calculated due to its separation from that current point. The weighting employed in the database is $\frac{1}{\text{separation}^2}$ which produces the most 'realistic', and statistically accurate, approximation to a topographic surface from data distributions which are typical within the map database.

Another factor in this averaging is the number of points included in each interpolated value. The lower this is (three or fewer points) the more jagged the resultant surface, because it includes discontinuities created as the search switches from one data point to another. But, if only a few points are available on the real data map, and a large number of points are specified (five or more), the distances over which points may affect one another becomes unrealistic and the resultant surfaces unnatural. In general, somewhere between three and six points should be used, with more points included if more real data are available. As an empirical rule - five points if 25% real data cover is achieved and three points for less than 10% original data coverage.

8.1.2 Reliability / proximity index

To provide an indication of the trust a user can place in a particular interpolated point the interpolation programs have been designed to generate a separate map of a segment's proximity to real data. This map is produced by mapping a pre-calculated, centred matrix of annular proximity fields on to the data distribution map for each data point. The best, i.e. closest, values are selected from the map and matrix at each element. If, when the proximity map is complete, a segment has no score no value is interpolated for that segment of the output map.

In the datastore as it is currently interpolated a standard set of proximity ranges are used, these are: 100, 200, 400 and 800 metres of separation. But, for special purposes the programs allow any set of up to six ranges to be specified. The maximum proximity radius catered for is 1 kilometre. Of course, to allow a 100% coverage by the interpolated map the matrixing may be switched off.

8.1.3 Ordinal data

A special program, PXTPOL, is provided to cater for the case when the number of included points is equal to one. In this event mutually exclusive Thiessen, or Veroni, polygons are created around integer data points. These mark out the area which is closer to that data point than any other in the set.

Such a map, called a Proximal map by Fisher in his SYMAP package (1968), provides an indication of the likelihood of finding the integer indicated by the interpolation in reality. This prediction is suited to the numeric coding employed in the database. But it must be treated with some suspicion as a tool for geological interpretation, because it is dependent on implicit underlying linear characteristics for the initial data. This linearity cannot be promised for the numeric equivalents used in the database.

The final maps, therefore, should be treated as a statistical prediction, based on spatial information, rather than a real parameter map. In areas of high data density this estimate will be fairly reliable, indicating streams, terraces, hillsides etc.. But the at distance use, because of the discontinuous nature of geological structures, will have a greater decrease in reliability with distance from actual data than the interpolated linear data maps, such as height O.D..

8.1.4 The use of barriers to interpolation

The barrier map facility in the interpolation routines allows the user to 'mask off' areas of the map from the data spreading, leaving blank regions. The effect is identical to masking areas of a photograph during printing. It enables several independent images to be exposed on the paper with no cross-contamination.

There are several possible uses for this effect. They centre around the creation of greater realism in the derivation process. For instance, if a particular parameter only has relevance when a certain lithology is present or if a real value is being interpolated for a particular stratigraphy which is limited to genetically disparate bodies, the points from one body must not be used to calculate the values in another body. In these cases a more realistic final map will be obtained if masks are made for each body of sediment and the separately interpolated maps superimposed to give the final result. Another example of the use of this technique is the calculation of parameters on either side of a fault, although in this case the age of the last movement must be carefully considered.

One final use for a barrier map is as an existence table to increase the efficiency of data retrieval and processing programs. Such programs are designed to 'look' at the barrier map and skip the processing of the segments where it indicates that the required data are not available, or required.

8.2 Combine / Scatter / Integerising

The two image processing programs, COMBINE and SCATTER, allow, with the custom combination facility, almost limitless manipulation of sets of maps. COMBINE may either be used with its standard functions and a large number of operations to achieve the required goal or simple FORTRAN subroutines can be added to the main program by the user to give the result in one operation. This latter approach is recommended for its efficiency and simplicity. Some examples of the many possible operations are given in Figure 8.1. It is well worth while bearing in mind the possibility of the user specifying his own lookup tables by pre-defining arrays in DATA statements and the use of the many sophisticated mathematical functions provided in FORTRAN.

The SCATTER program has two uses, firstly it displays

Figure 8.1 Examples of custom sections for COMBINE

```
C  selection of the greatest value between two maps

      X = A
      IF (A.LT.0·0) RETURN
      IF (B.GT.A) X = B
```

```
C  conversion of general lithologies into single code
C  Glacial values

      X = A
      IF (A.EQ.2) X = 6
      IF (A.EQ.3.OR.A.EQ.4) X = 7
```

```
C  lookup of single strat. codes for specific strats.

      DIMENSION I(47)
      DATA I/6*8,9,3*0,7*6,7,22*0,0,0,2,3,4,5,0/
      IF (A.EQ.0·0) RETURN
      X = FLOAT(I(IFIX(A)+1))
```

```
C  lookup for broad lithologies from specific codes

      DIMENSION I(94)
      DATA I/0,4*4,3*3,2*2,3*3,2,3,2*2,2*3,11*0,4*4,3,2,3,4,
     13,0,2*3,2*2,2*3,9*2,5*0,4,2,3*1,2*2,4*3,1,2*3,4*1,9*0,
     15*1/
      IF (A.EQ.0) RETURN
      X = I(IFIX(A))
```

```
C  reduces the symbol shading of isopleth maps to symbols
C  separated by C1 spaces

      X = 0·0
      I = I+1
      IF (I.NE.IFIX(C1)) RETURN
      I = 0
      X = A
```

the scalar distribution of the data and secondly it facilitates the division of that information into an integer form. Thus the user may both control and change the type of map being produced and analyse it, via the histogram and CHORO display.

Reference has already been made to the problems and advantages of the four significant figure format of the map files. In addition, there exists a limiting factor to the control of the subdivision of classes in SCATTER. The screen image is set at 80 columns and thus the class boundaries must be set at one of these values. In general this does not cause problems but, if fine detail over a small range is required, the COMBINE program may be employed to limit the spread of values, using options 8 and 9, prior to the integerising of the map. (ref. p 148)

For display purposes the CHORO output has so many levels of colour and/or shade that the plotting of maps of real data in an understandable form is possible. The DRAW routine, with its six shade levels, must be used more carefully.

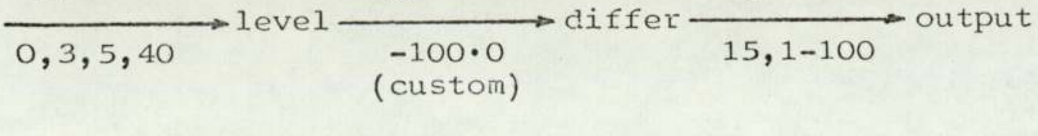
Proximity maps are obligatory for the COMBINE routine. These are compared by the program and the lowest reliability value specified is assigned to each segment of the output map. This can lead to some problems where full data coverage is required, for instance where one map is much less significant than another or if a large number of disparate data sets are being combined. At best a uniform poor proximity is achieved whilst in the worst case large areas become effectively devoid of information. A partial solution to this problem is given in Appendix 3. But, due to the complexity of the necessary electives, the permanent implementation of this 'proximity correction' cannot be justified.

8.3 Array Processing

The use of the EXTRACT, COMBINE, SCATTER, EXTPOL, PXTPOL and REPLACE programs in concert, to bring about changes or enhancements to the datastore can be planned as a set of discrete logical operations. The information is transferred from map file to map file via 'transformations' which are the system programs. This structure generates two types of operational blocks which allow a MACRO SCALE PROGRAMMING to be derived and thus facilitate operation planning.

A schematic format which suits this macro scale programming describes the map files by mnemonic titles and represents the functions, or programs, by arrows titled with the first letter(s) of the function name, i.e. E,C,S,EX,PX,R. An additional note, specifying the particular electives or describing the custom subroutine employed, may be added. Figure 8.2 shows the instructions necessary to extract the relevant data, calculate the deviation of the stored bedrock surface from the 100m contour and present the result as a map for CHORO.

Figure 8.2 example of a macro-scale program



8.4 Example One, Extrapolation Into A Map

8.4.1 Definition of purpose

The requirement for a fuller coverage of the mapped area than that achieved by the site investigation data was mentioned in the initial definition of the project. This data spreading is obviously a complicated task, if the interpolation of the soils and rocks is to be geologically related. The advantage of carrying out an initial, fully computerised interpolation is the relatively simple production of a three dimensional, multiparametric model of the area. This will provide an indication of future areas and methods of study in the construction of a more interactive and detailed 'map'. The creation of such a program, in a usable form, is beyond the time and facilities available to the current project.

Some broad constraints for the interpolated map must be defined:

The information is to be worst case but not extreme, to prevent the over-influence of single data values which may well be spurious.

A consideration of the geology of the area demands that each stratigraphic horizon should be treated independently (at least as far as its lithology and geotechnical properties are concerned).

The interpolated information must be distinguishable from the original and synthesised data, with some indication of its reliability.

8.4.2 The three dimensional model of the stratigraphy

To enable the various geological and geotechnical parameters to be interpolated for only the areas in which their stratigraphy exists a three dimensional model for the structure of the area, at a stratigraphic level, must be created. This model should take account of all the available information in the datastore. The following synthesis was used to employ all of the relevant data to produce a worst case model.

8.4.2.1 theory of the process

Only the bedrock is present globally* in the map area.

We have real boundaries recorded between geological layers.

We have values recorded for the bases of boreholes terminating inside a certain stratigraphy.

These 'lowest recorded' data can be used to provide best case, i.e. minimum depth, extra detail for the real boundary data, when these values are below those calculated from the real data.

The adjacent improved lowest limit surfaces thus obtained can be compared to find the lenses bounded by them which will contain a particular stratigraphy.

If a predominantly erosive regime is assumed it can be seen that, working down from the top surface (the present day topographic), and successively 'cutting off' the next surface down, where it protrudes above the younger surface, will give both the next 'younger' surface, for a repeat of the operation, and an existence map for the stratigraphy associated with the older layer.

8.4.2.2 practice

Extract the lowest levels of the Made, Recent and Glacial (i.e. 20,0,10) stratigraphy. Combine and interpolate

---

* In databasing the globe is the datastore, the world is all human knowledge.

with no proximity limits.

Extract the tops of the 0,10,40 stratigraphies and interpolate with no proximity limits.

Compare the base of the 20,0,10 with the top of 0,10,40 respectively.

The lower of these two values is chosen as being a 'better' approximation to the real surface. However, it is only a best case in the areas where real data is not present for the lower surface.

The residual best case surfaces are successively compared with the surface stratigraphically below them and new upper limit surfaces for the older stratigraphies produced.

FILES USED:	COD	topographic O.S. 1:10,000
	TR2	best base of the Made Ground
	TR0	best base of the Recent
	TR1	best base of the Glacial
	RR2,0,1	barrier or existence maps
	BB2,0,1	boundary height maps

The space below the 'base of Glacial' is implicitly bedrock. The function used in the COMBINE program is option 5.

OPERATIONS:

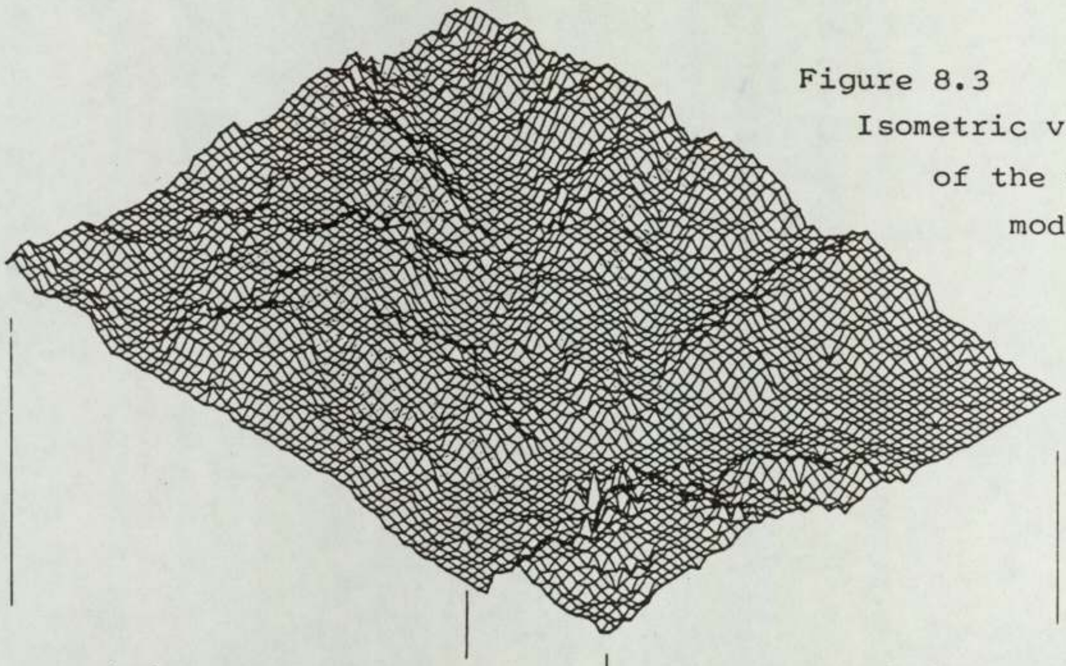
younger <u>vs</u>	older	→	lowest	+	barrier map
TR2	COD	→	BB2		RR2
TR0	BB2	→	BB0		RR0
TR1	BB0	→	BB1		RR1

The existence map for the bedrock, RR4, is naturally always equal to one. BB4 is set at BB1 minus one metre.

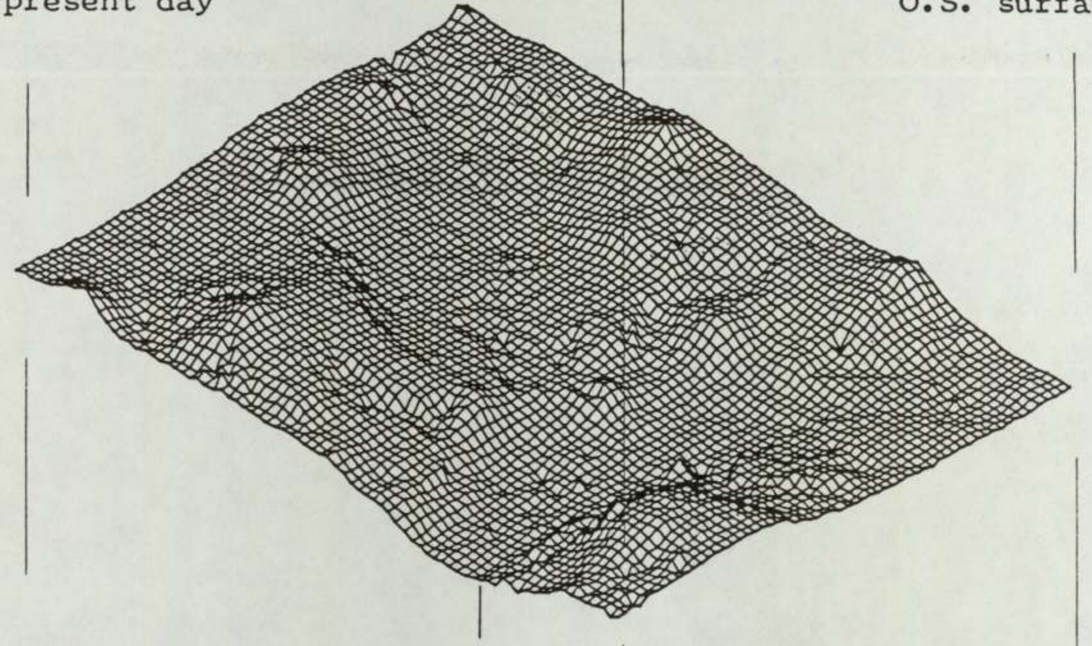
8.4.3 Method and cautions of spreading data in the model

The 'filling in' of the three dimensional spaces represented by the files BB2,0,1 and RR2,0,1 and shown in Figure 8.3

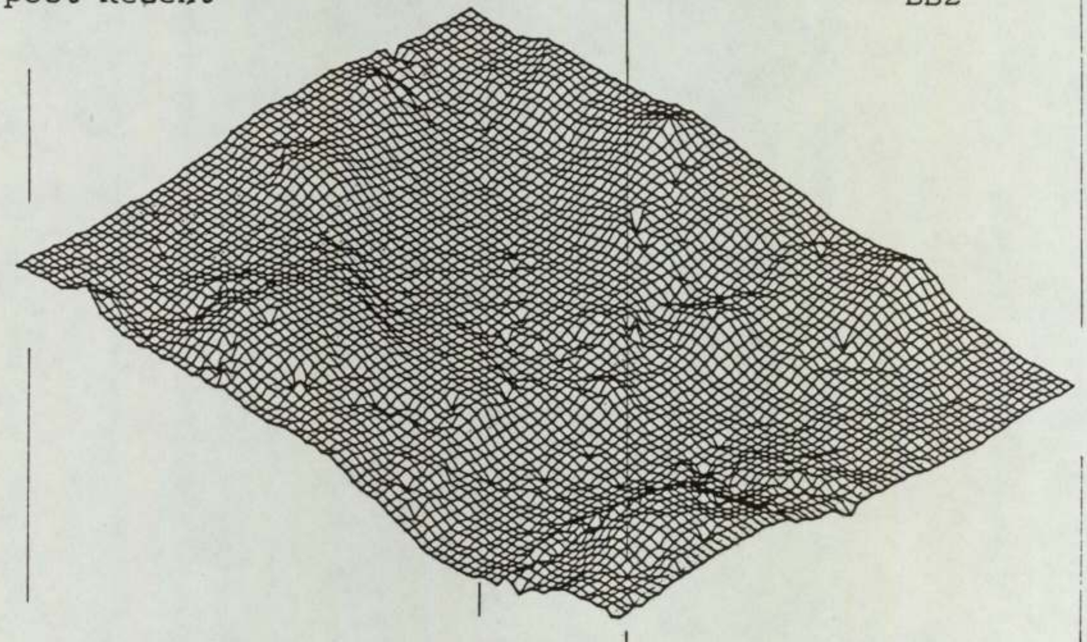
Figure 8.3 Isometric views of the 3-D model



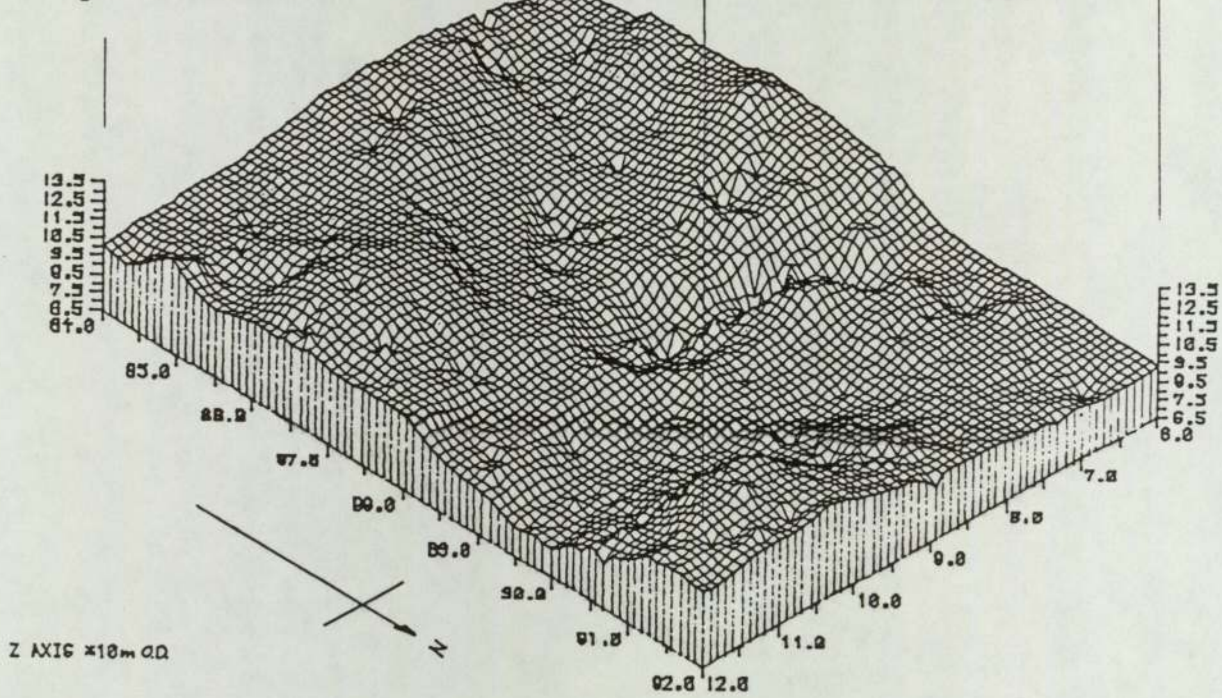
present day O.S. surface



post Recent BB2



post Glacial BBO



rockhead BB1

is not complicated, but it is tedious. The process entails EXTRACTing the relevant dataset from the old (RAW) datastore and interpolating it, with a proximity map and the appropriate barrier, to a map which will be included in a new datastore by a REPLACE program. Three REPLACES are required, for the header, geological and engineering data respectively. A basic nomenclature is employed:

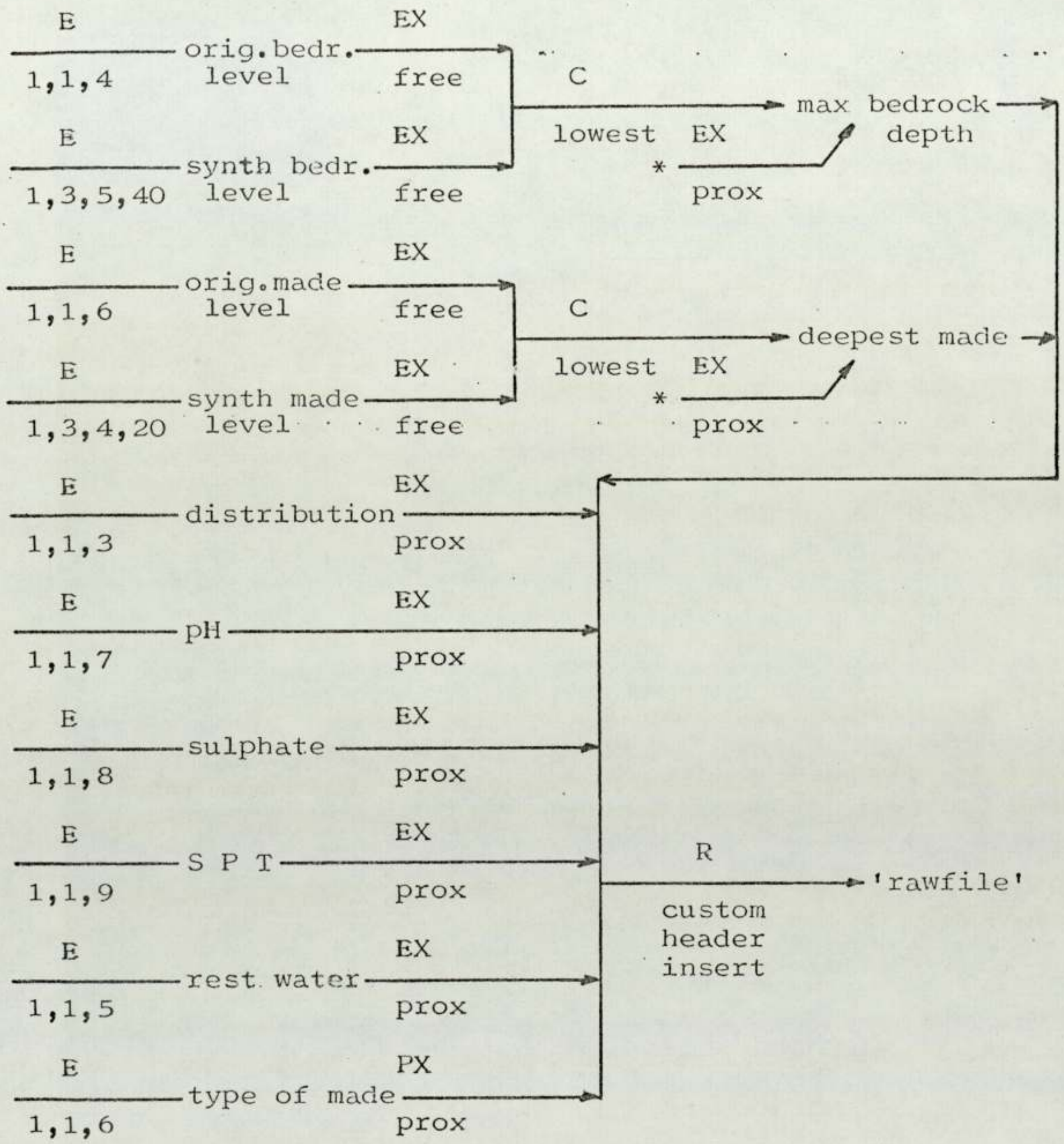
R-	real data map
I-	integer proximity map
MX-	extreme value map
T-	interpolated map of data
IT-	proximity map for interpolated data

The file names are chosen to describe the contents as far as is possible.

Figures 8.4, 8.5 and 8.6 show the three stages in the generation of the current FUL- datastore. File names ending in FILE are the total direct access filestore and its index.

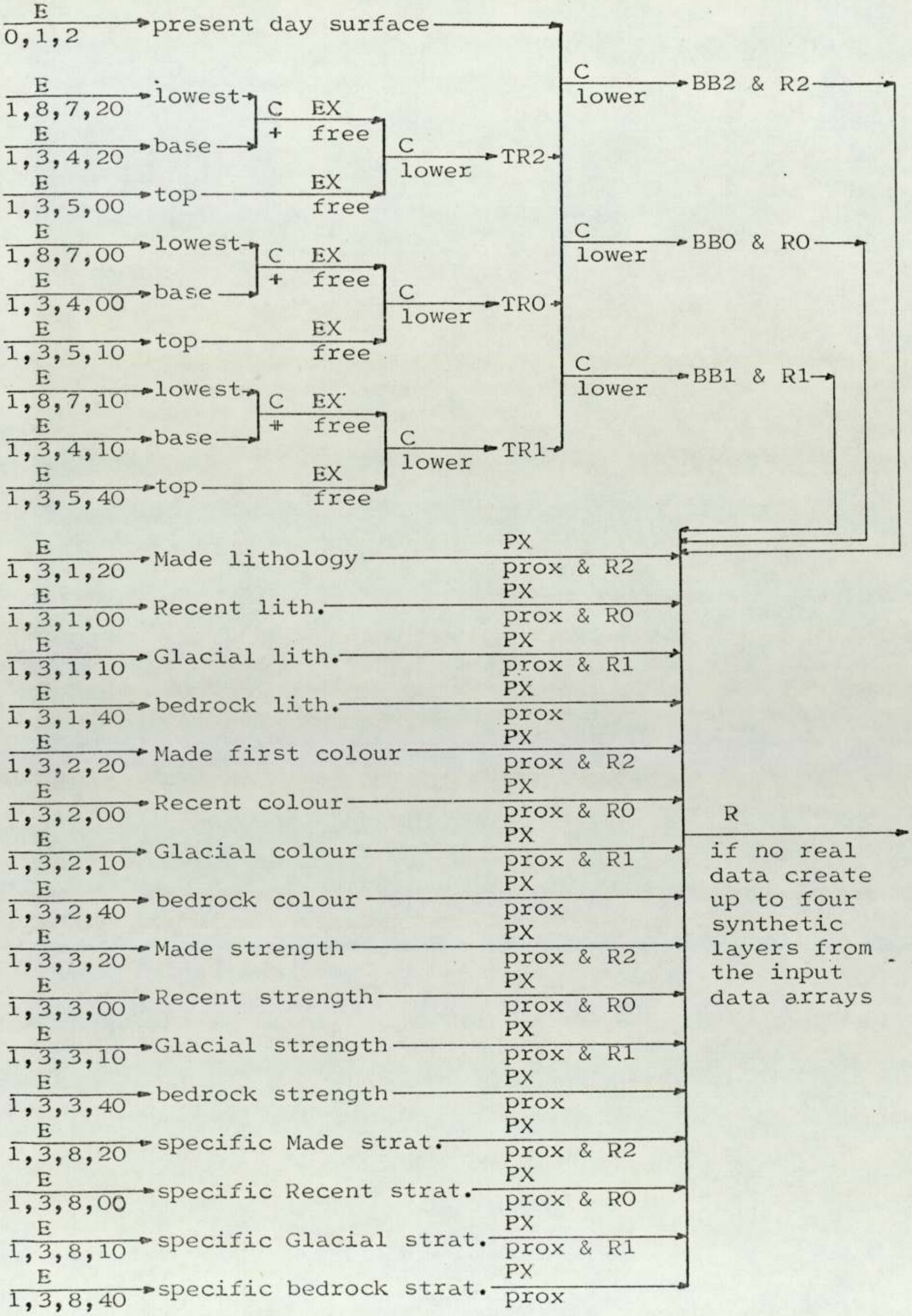
A number of criticisms can be levelled against this extrapolation technique. The inherent failings of the interpolation routines have already been illustrated. Because the information was derived for interpolation at a stratigraphic level it has little significance below that level of detail. This is not important for the engineering data since it was already stored at a stratigraphic level. However, a cross SECTION drawn through the full datastore indicates the potential for improvement in the geological analysis. This could be achieved by a repetition of the existing technique at a lithological level, or an interactive sketching routine. But it should be noted that the initial site investigation information lacks resolution below a stratigraphic subdivision, making conclusions less reliable...

Figure 8.4 Interpolation of the Header Data



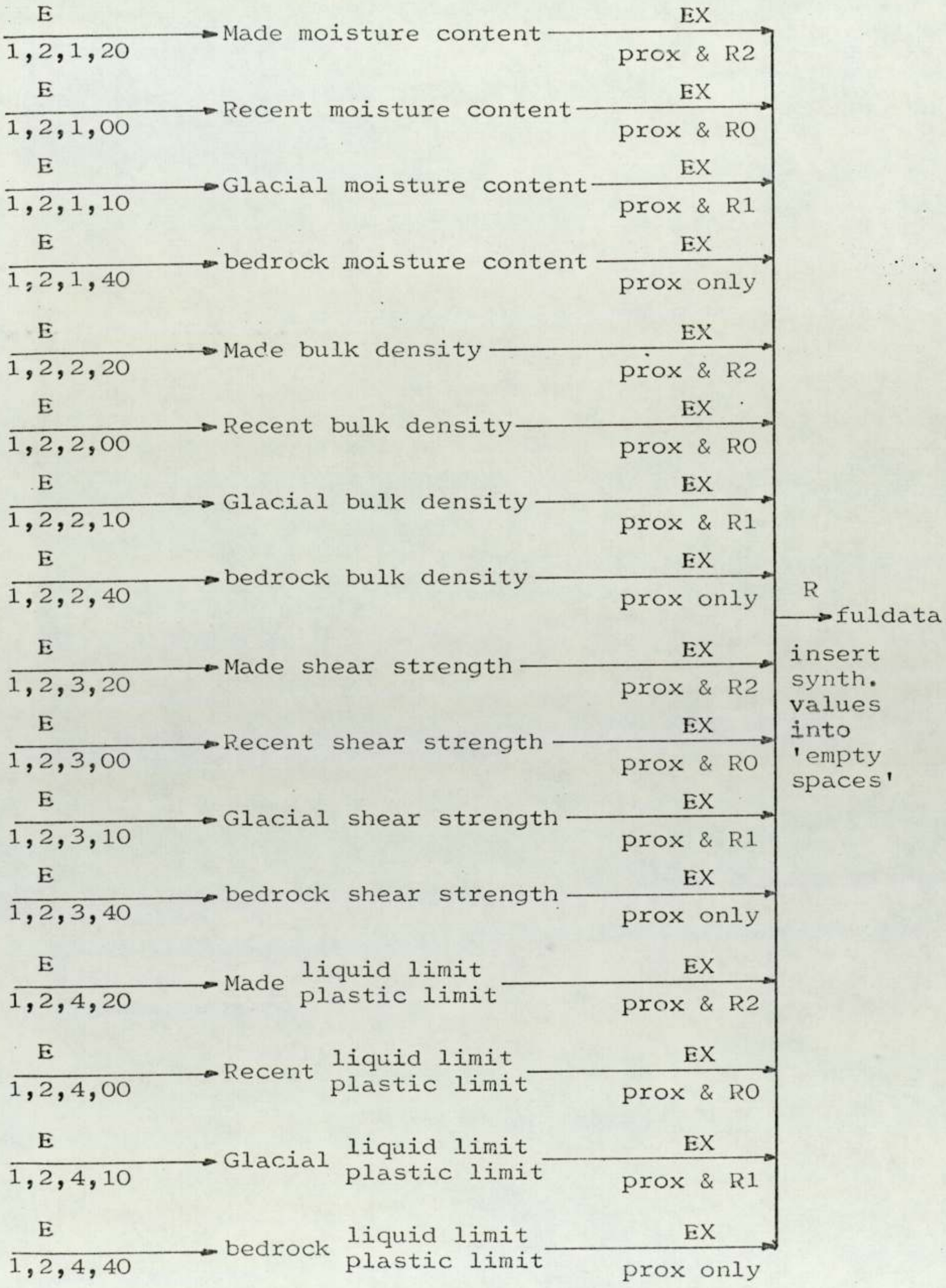
* The processing of the proximity maps for each parameter is implicit in the macro scale programs. The proximity for the maximum depth bedrock and Made is calculated from the sum of the source map distributions.

Figure 8.5 Creation of Interpolated Geological layers



The geological properties are the commonest codes for each parameter for the specified general stratigraphy.

Figure 8.6 Creation of Interpolated Engineering Records



The values for the various parameters are the worst average for each general stratigraphy.

The header data (specifically the water, worst bedrock and worst Made) are not derived or corrected from the interpolated values in the stratigraphic model. This is an attempt to provide a link back to the original information as a further index of reliability. The stored values are interpolated only from original, or 'hard', unsynthesised data.

Although the header, geological layer and engineering information were interpolated with discrete proximity maps for every set of data the geological layer file uses only the overall borehole distribution as its final proximity map. This is done mainly to save space, otherwise the layer file would be by far the largest in the store. Because almost every layer has a lithology, primary colour, stratigraphy and qualitative strength there are very few differences between the overall distribution of the boreholes and that for the specific geological parameters. The proximity maps are naturally altered to take note of the few extra data absences during EXTRACT operations.

If the geological data are studied it will be noticed that the lithological parameter is poorly interpolated. This is because the numeric codes used are direct equivalents of verbal descriptions and these have evolved with time. The different codes generate discrete Thiessen polygons. This leads to sedimentary bodies that, although they are quite clearly of uniform composition, have an appearance which is scrambled by the interpolation. If the source information was homogeneous, or made so either by calculation or coding, this problem would not occur. Incidentally, the use of proportional coding for the lithology and colour would be of advantage in this particular operation, generating a more natural variation, even though the interpolation would involve many more operations and the result occupy more storage space.

8.5 Drafting Operations

8.5.1 CHORO

Two programs (CHORO and DRAW) produce maps. CHORO is less complex but only of use in the creation of isopleth maps. The screen and printed maps are different in that the VDU image is made up of coloured blocks whilst the line printer copy consists of pairs of printed ASCII symbols. This difference is not important when each class displayed represents a discrete entity, such as a stratigraphy or an archive type. But, when shading is required, there is much more difficulty in distinguishing between 15 printed symbols than 15 homogeneous shades, especially if these shades are selected to range between two complimentary colours.

The use of different VLTs is well worth investigation. For instance, a topographic surface looks far more 'real', and understandable, if it is shaded in the standard blue/green/brown/purple/white of the atlas map than a scale with colour tints from red to blue. However, this second scale is ideal for an indication of pH.

8.5.2 DRAW

The DRAW program produces similar output on a number of devices by line drawing. It is designed to plot isoline maps and isometric views but it can also produce flexible symbolic maps. Many map files may be plotted at one time but if this is to be done, care should be taken to avoid a program crash, because restarting the long creation of a complex output map is frustrating and expensive.

Single maps may be enhanced by the careful choice of colour and the addition of a shaded proximity map. These maps are processed by COMBINE to transform the negative proximity and positive real retrieved map into the 1-6 class values required by the symbolic plotter. Contours can be enhanced by plotting the same map between several vertical limits with each set of contours in a different shade. For example, giving a blue to red trend for the contours of an isopachyte map.

If several maps are plotted on the same picture, to show their inter-relationships, colour may be used to differentiate between the maps. Time is well spent in altering the vertical spacing of the contours of the maps to increase the clarity of the result.

The symbol plotter may simply be used for shading areas of a map but it also provides control over the symbol's colour, type and size. This gives great flexibility to the display, allowing a great deal of information to be shown on one picture. Examples are:

use of the WPR'72 alphanumeric lithology symbols for soil maps

plotting line shaded reliability maps

addition of stippled hill shading to contour maps

use of the size of a symbol to indicate another parameter, such as colour, stratigraphy and strength in the colour, type and size respectively.

For this last operation the COMBINE program, option 7, is used to break a symbol size map (strength = size) into, up to four, separate maps of colour and/or texture. These are then submitted to the DRAW program and plotted in one run for the 'n' colours and/or textures.

Isometric plots are often more convenient than contour maps. The program gives control over the vertical exaggeration of the diagram. This enables either realistic pictures of parameters, such as the relief of a particular age of surface, or enhanced representations of engineering parameters, such as the Standard Penetration Test results.

The pseudocode generator and the PSOUT program provide the facility to make a 'master' map and run off a number of copies. Another use for PSOUT could be to make up a composite picture from several pseudocode files, using the ability to move the starting point (or bottom left hand corner) of the image. Some examples are the plotting of a set of isometrics from all four view angles or more complex superimposition than can be achieved with the DRAW routine alone.

8.6 Handling Sets Of Parameters

Morin (1979) divides the type of maps derived from a geotechnical databank into three classes: Documentation, Derived and 'Synthesis'. Each class represents a more sophisticated manipulation of the existing information. For this project these three classes have been adopted, with slightly altered definitions to allow for the use of interpolation procedures which were not considered by Morin.

8.6.1 Documentation

Single parameters read directly from the datastore with little or no intermediate processing. Examples of these maps are:

Engineering parameter charts. Real numbers indicating actual and interpolated values for the stored results. These can often be discontinuous and the best mode for representation is, at the discretion of the user, the isoline or isopleth map.

Geological information, such as qualitative strength, is best represented on isopleth maps.

Archive data, requires the use of located symbol maps.

The topography and palaeotopography are well suited to isometric display and if measurable values are required the isoline map is also available.

It should be noted that some parameters may be further decoded from the form in which they are held in the database. The sulphate concentration can be split up into water and soil readings (by COMBINE, using the information provided in the previous chapter) and these may then be interpolated separately to give a more realistic representation on two discrete maps.

8.6.2 Derived

The use of statistical, mathematical or logical combinations of sets of data to represent some parameter more fully or to create some value not actually held in the datastore. Quite advanced interpretations may be carried out in this manner. The structure of the map files and the COMBINE program in this database were designed specifically with this type of analysis in mind.

Examples of the use of derived maps for the engineering data are the calculation of the degree of saturation from the bulk density and moisture content or the Consistency Index for cohesive soils. Useful maps may often be created by adding geological parameters, such as the fines content, to engineering information. Morin (1979) mentions the possibility of making up maps of indices which require some specific design parameters, such as the size and weight of the proposed structure.

Geological manipulations are usually logical with a statistical justification which is based on the behaviour of the interpolation and extraction programs as well as that of the original geology. The best way to demonstrate geological processing is to give three examples, all of which are used in the presentation maps which make up Appendix 7.

8.6.2.1 isopachyte maps for the various stratigraphies

Illustrated in Figure 8.7, this is a simple matter of subtracting the various levels, recorded and extrapolated, in the datastore. It uses no custom COMBINE operations. In the printed set of maps these thicknesses are presented as the variation in the sizes of symbols for the sediment type by further COMBINE manipulations, as described on page 180.

8.6.2.2 simple lithostratigraphical drift map

The custom facility is employed three times here. Firstly to translate the lithology into stratigraphic codes. Secondly to remove some 'noise' in the interpolations by cutting off thin, or unreliable, layers. Thirdly to choose the recent deposits over the glacial in accordance with the law of superposition. These are shown in Figure 8.8.

Clearly the simplistic view of the subdivision of the Glacial strata herein is not in accord with the Geological Survey's classification of the area. However, the resolution of the interpolated lithology will allow no more sophisticated division at present. It must also be remembered that the database 'spreading' took the commonest lithology from each real 'well'. This preselection 'filters' one typical soil type for each stratigraphy in that segment. The

Figure 8.7 Program for the stratigraphic Isopachyte Maps

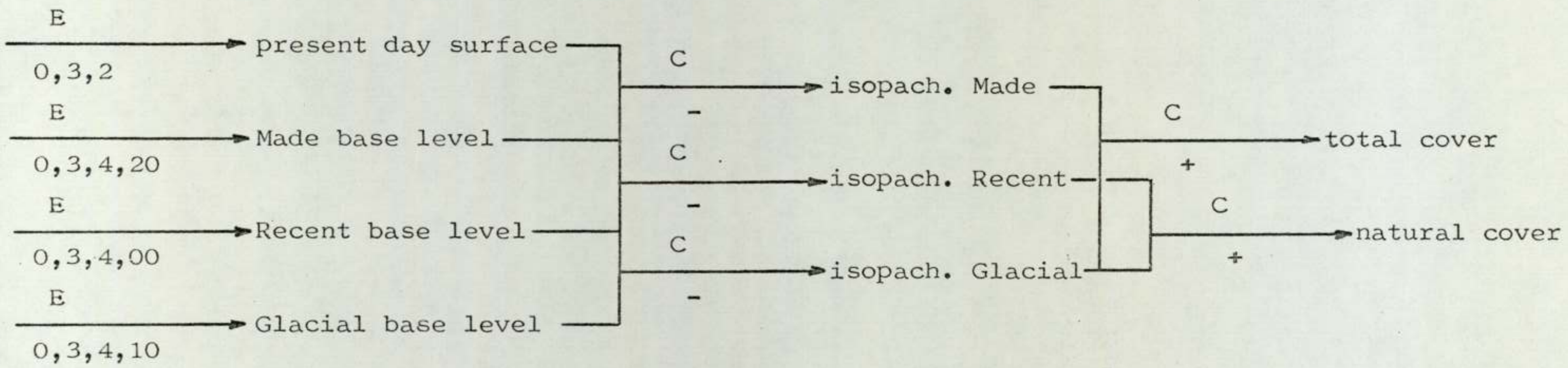
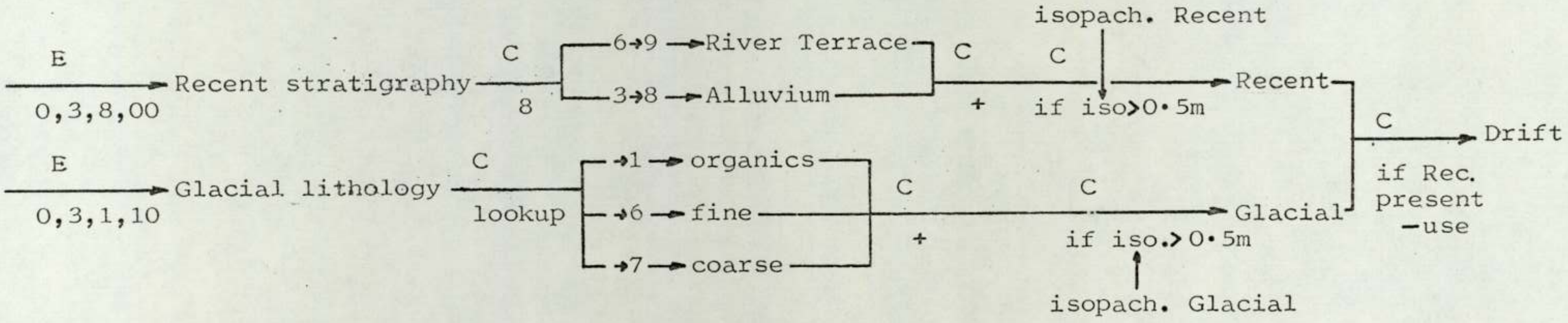


Figure 8.8 Program for the preparation of a simple Drift Map overlay for the Bedrock map



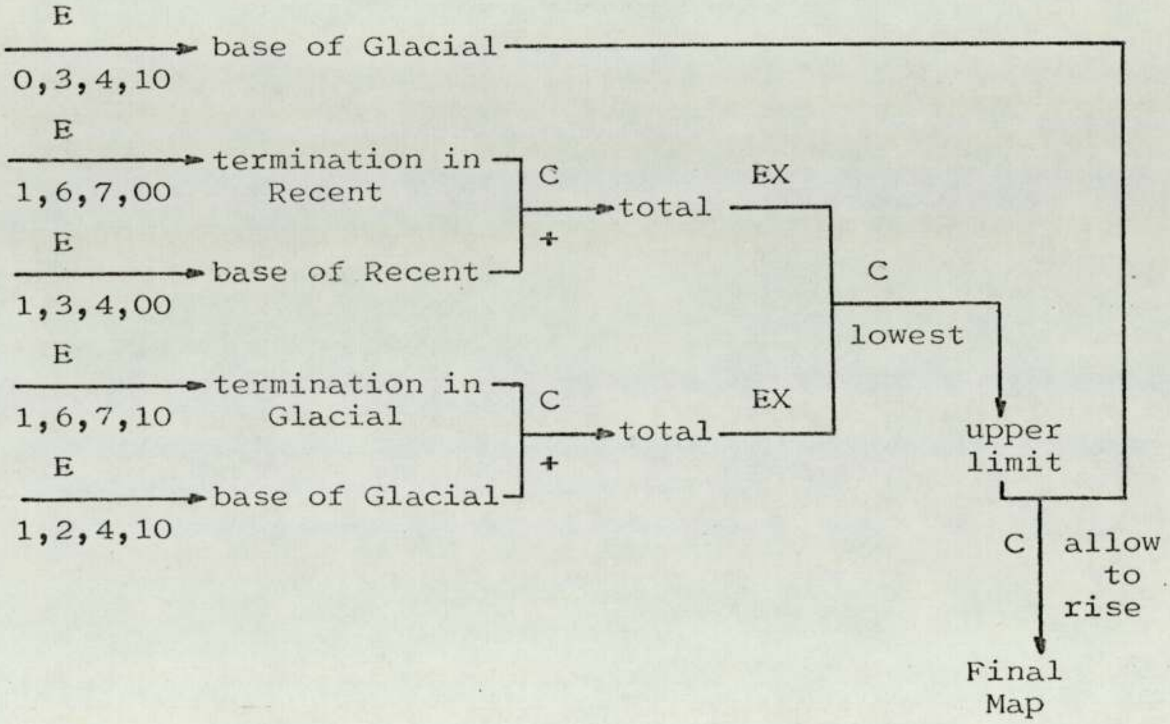
map generated should thus be looked at in that light, as a a fast-to-produce lithostratigraphical presentation exercise rather than any challenge to conventional geological maps.

Inclusion of engineering parameters in this classification, as suggested by Horton (1974,1975), to group the soils according to typical test results could produce much better results, as long as the scatter of the samples was minimised by employing as many parameters as possible.

8.6.2.3 a best rockhead relief map

The theory behind the enhancement of the stored surface shown in Figure 8.9 is to reduce the effect of the Present Day and Made surfaces on the rockhead level.

Figure 8.9 Program for the preparation of a 'best' Rockhead Surface



A surface is derived from the measured bedrock boundary found in segments and from the lowest levels proved in segments which do not reach bedrock. This defines the upper limit surface of the bedrock. The stored 'worst case' surface is then allowed to 'rise' against this surface, thus eliminating many of the troughs caused by the present day and pre-Made ground topographies. A further extension which may improve the realism of the final surface is to remove the Recent surface effects as well. However, the control of the 'enhancing' surface would suffer with this further reduction in the number of points and, as the Recent deposits are never very thick, this would lead to only a small improvement.

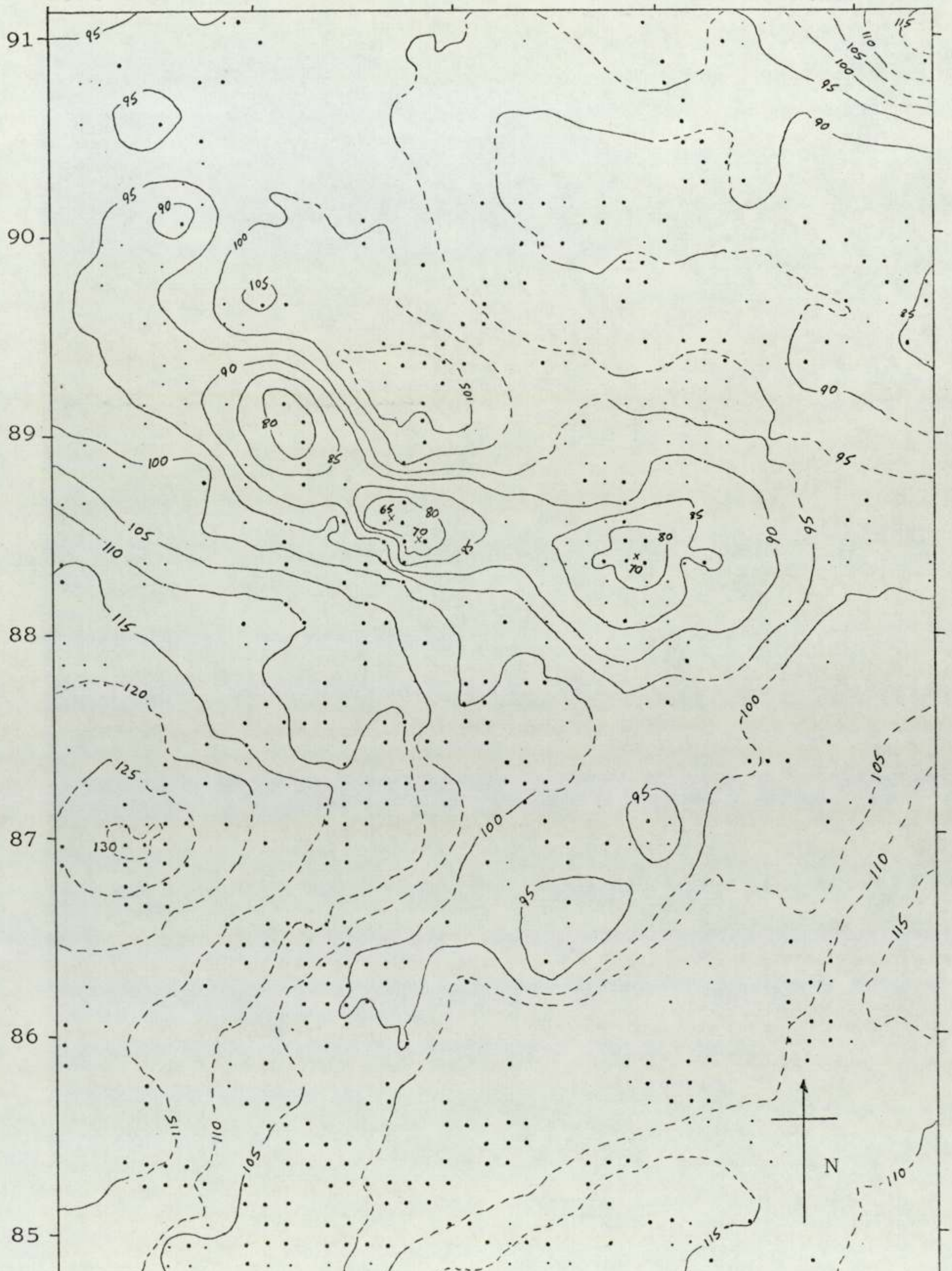
It must be remembered that in all this semi-intuitive manipulation of data from subtly different sources the main criterion of the final result is that it 'looks' more geomorphically correct than other attempts. At this point it should be noted that the buried topography is, from its lowest valley to its highest 'peak', shorter than one of the office blocks in the city centre. The isometric plot is excellent for viewing the general morphology but when it is used realistically, with no vertical exaggeration, no features are discernable on the topographic surfaces. In this case the familiar contour map plus the world knowledge, or preconceptions of form, of the viewer are invaluable.

The rockhead map shown in Figure 8.10 is drawn from the 'best rockhead' surface. The borehole locations are taken directly from a DRAW map of the overall segment data distribution. The DRAWn contours are made broken or solid depending on the absence or presence of glacial cover respectively. The presence of glacial deposits is taken to indicate that the rockhead topography in that area is Glacial in age. This existence information was derived from a scaled-down CHORO plot of the glacial barrier map for the map database model.

Maps of the model's 'worst case' rockhead and the

Figure 8.10 Rockhead Relief for an area of the Map

borehole · to bedrock • level OD ---- sub-Glacial ——



present day surface were DRAWn out and COMBINE/CHORO maps of the difference between these surfaces and the 'best' were created. All of these were employed in an analysis of the 'best rockhead' surface.

The main feature is the proto-Tame valley, running from 0690 to 090885. To the east of the latter point no more real data exists on which to base a firm estimate of the course or depth of this valley. It therefore fades out on the map, even though it must exist in this area.

A strong river terrace is developed in the area around 085850 to 1088. This dips slightly to the northeast. It is a Recent feature of the Rea valley but it has subsequently been both excavated and buried by man, especially in the region of the Adderley refuse tip (ref. Appendix 7, O.S. map).

To the west of this terrace a slight trough is present. This is Glacial in age and may represent a contemporary stream in this area. The River Rea itself ran to the west of the mapped area in glacial times, Pickering (1957).

The CHORO difference maps reveal a large discrepancy in the 'worst' and 'best' rockheads in the Gravelly Hill region. This is attributed to the complexity and confusion of the level information and Made deposits in that area, caused by many generations of transport route embankments and excavations. The 'best' map, however, manages to eliminate much of the confusion and shows only the increase in slope of the north bank of the Recent Tame valley.

8.6.3 Synthesis

Either the presentation of existing information in advanced combinations which are of relevance to specific groups of individuals or the interpretation of parameter sets to produce new facts. One example of this synthesis is the aforementioned use of engineering information to enhance the drift map of the area. Others, suggested by Morin (1979), are maps of relative cost of excavation of a cubic metre of soil or a first appraisal-of-sites map with development difficulty on a scale of one to ten, derived from simple characteristics such as cover, water depth and soil type.

8.6.3.1 the summary index or 'hazard' map

In the definition of the current project there exists a requirement for some form of synthetic summary map for the stored area. This map is intended to provide a !quick glance' index to the usability of a particular site. This will hopefully be of value to most of the users of the map database system as a starting point for an investigation.

The generation of a single map, rather than a set of ten or so discrete sheets, is an attempt to make the earliest stages of site analysis simpler and more accessible to users by pre-interpretation of the critical factors.

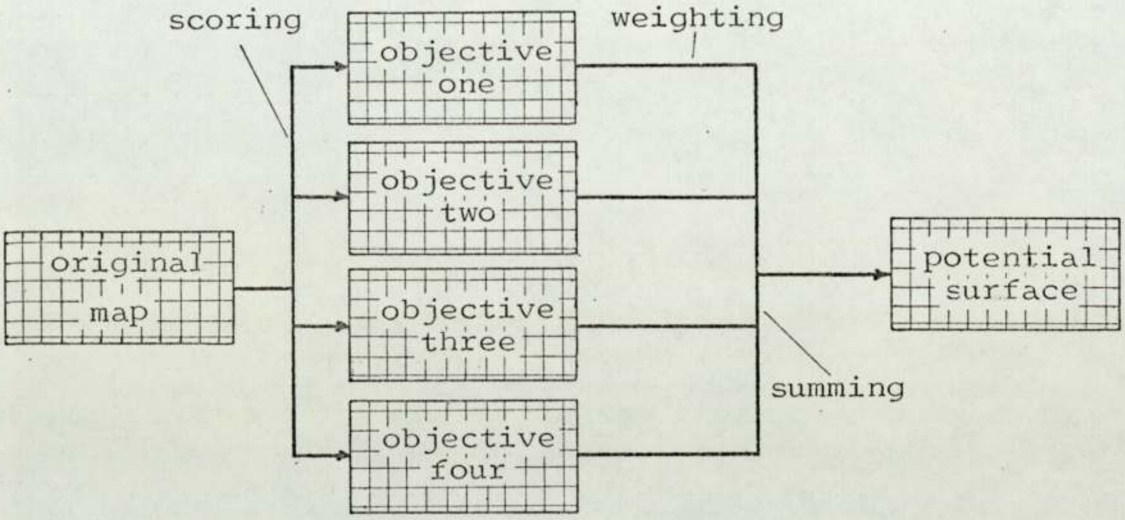
In the description of an entity, Sokal and Rohlf (1970) found that as more characteristics were employed in the definition the noise, or misidentification of the type, decreased in proportion to the number of identifiers used. If this observation is applied to a synthesis map, a summation of many (individually quite unreliable) descriptive parameters could produce a fair overall indication of the 'quality' of the area for, in this case, engineering use.

8.6.3.2 potential surface analysis

PSA is a technique for: 'systematically assessing the potential of an area to accommodate a particular type of development or land use and for presenting the findings in an easily understood way which can be manipulated to show the effects of changing assumptions and objectives.' Zetter

(1974). Figure 8.11 shows this sieve mapping technique in operation.

Figure 8.11 Creation of a Potential surface



The factors which make up the map are scored for a particular 'objective' - this may be anything from a highly qualitative estimate to precise quantities of specified variables. Information is collected by suitable measurement techniques from the study area on a regular grid. The various objectives, or parameters, are weighted with respect to one another. The sums of the weighted values for each grid element make up the final potential surface.

An important point to bear in mind when using this technique is that real and integer objectives are treated identically. This can produce round-off errors in the case of a number of real values and more importantly, can lead to integers being treated as real numbers. This latter problem becomes acute if the potential surface is displayed by a continuous representation, such as an isometric plot, or, if the surface is re-integerised (classified), after large weighting factors have been used. However, when a number of parameters, with reasonably similar weightings, are used individual breaks between the classes of a particular integer become less important in comparison with the whole score distribution and while (especially during re-classification) the problem should not be ignored, it can at least be tolerated.

8.6.3.3 practical generation of the map

The initial parameters for the hazard map of the area are the various sets of information held in the datastore. For most of these data predefined classifications may be found in WPR'72 or IAEG/UNESCO (1976). These divisions almost always have five classes. Other parameters which are included will also have to be divided into five classes to simplify the relative weighting of the maps.

Standard definitions are employed for most of the engineering parameters. The thicknesses of Made Ground and total cover, combined from the worst average and maximum depths, were divided at approximate foundation design changes. The consistency index was classified by division of its distribution histogram, to give similar class score proportions to the other engineering parameters. Archive and the worst made type were given empirical scores depending on an estimate of the extra work required to make the area usable. The pH division was also empirical and was based on the distribution histogram.

Sulphate concentration was calculated separately for the water and the soil and then classified according to its effect on concrete by BS 4027, in Varey (1977). The results were combined onto one total sulphate map and share one proximity map. Qualitative strength is specified by WPR'72 descriptions and divided as such. Lithology is broken up by grain size and organic content. The distribution histograms and CHORO maps for all of the objective parameters for the hazard map are given in Appendix 6.

Weightings for the various parameters attempt to indicate both the relative importance of a parameter and the number of similar parameters which are in use. For instance, all of the cohesive soil parameters together have the same weight as the sulphate concentration or the pH but the thickness of the cover is considered more important than either of these. Table 8.1 indicates the parameters, their scores and their weightings.

Because the values held in the store are partly interpolated values the proximity index must be indicated in the final hazard map. If a simple worst case proximity selection were used the absence of any objective parameter, however unimportant, could delete a segment from the map. The proximity priority system mentioned earlier in this chapter was employed to ameliorate this effect. The minimum proximity factor given to each parameter is indicated in Table 8.1.

Summation of the scored, weighted maps presents a problem. If the values are simply added a low score highly weighted parameter will swamp a 'dangerous' high score with a low weighting. Some form of non-linear skew must be added to the distributions of the scores before they are weighted and added together. This skewing must be great enough to separate out the class five scores for the major maps but at the same time it must not be so large as to cause a major overflow of the four significant figures of the map files.

After trying out score^2, score^3 and 10^{score} a factor of score^4 was used. This produces a detailed distribution

Table 8.1 Classes and Weightings for the Hazard Map

type	1	val	2	val	3	val	4	val	5	weight fact.	corrn (px)	prop
isopach made meters	thin	2·5		5·0		7·5		10·0	thick	8	0	level
isopach bedr. meters	thin	2·5		5·0		10·0		20·0	thick	8	0	
total sulphate %	low	·2/·03	fair	·5/·12	mod.	1/·25	high	2/·5	v.high	6	-4	chem-istry
pH	neut.	6/8	fair	5/9	mod.	4/10	strong	3/11	v.str	6	-4	
archive	safe		fair		fault		tips		toxic	4		lith-ology
lithology	rocks		coarse		fine				peat	2	-2	
gen.stren. index	rock	11	stiff/ dense	13	firm/ loose	16	soft/ v.loose	18	peats	4	-3	strength
S P T 30cm^{-1}	v.dense	50	dense	30	mod.	10	loose	4	v.loose	4	-3	
shear kNm^{-2} strength	hard	150	stiff	75	firm	40	soft	20	v.soft	4	-3	
consistency index	brittle	-·005	pl.lim.	·005	plastic	·15	v.plast.	·5	liquid	2	-1	corroborative properties
plasticity index %	non pl.	1	slight	7	mod.	17	high	35	v.high	1	-1	
liquid limit %	lean	20	inter.	35	fat	50	v.fat	70	e.fat	1	-1	
saturation %	dry	25	wet	50	v.wet	80	h.sat.	95	sat.	1	-1	
unit weight kkgm^{-3}	v.high	2·2	high	1·9	mod.	1·7	low	1·4	v.low	1	-1	

the combination function is $\text{sum} = \text{sum} + \text{weight}*\text{data}^4$, corrn is the max. proximity effect histogram, a small proportion of overflows (which are automatically limited to a 9999 score) and a map with good detail in its features. The final map is divided into four classes at obvious break points in the distribution histogram, shown in Figure 8.12. These classes indicate definite levels in the relative difficulty of development in the environs of the City of Birmingham.

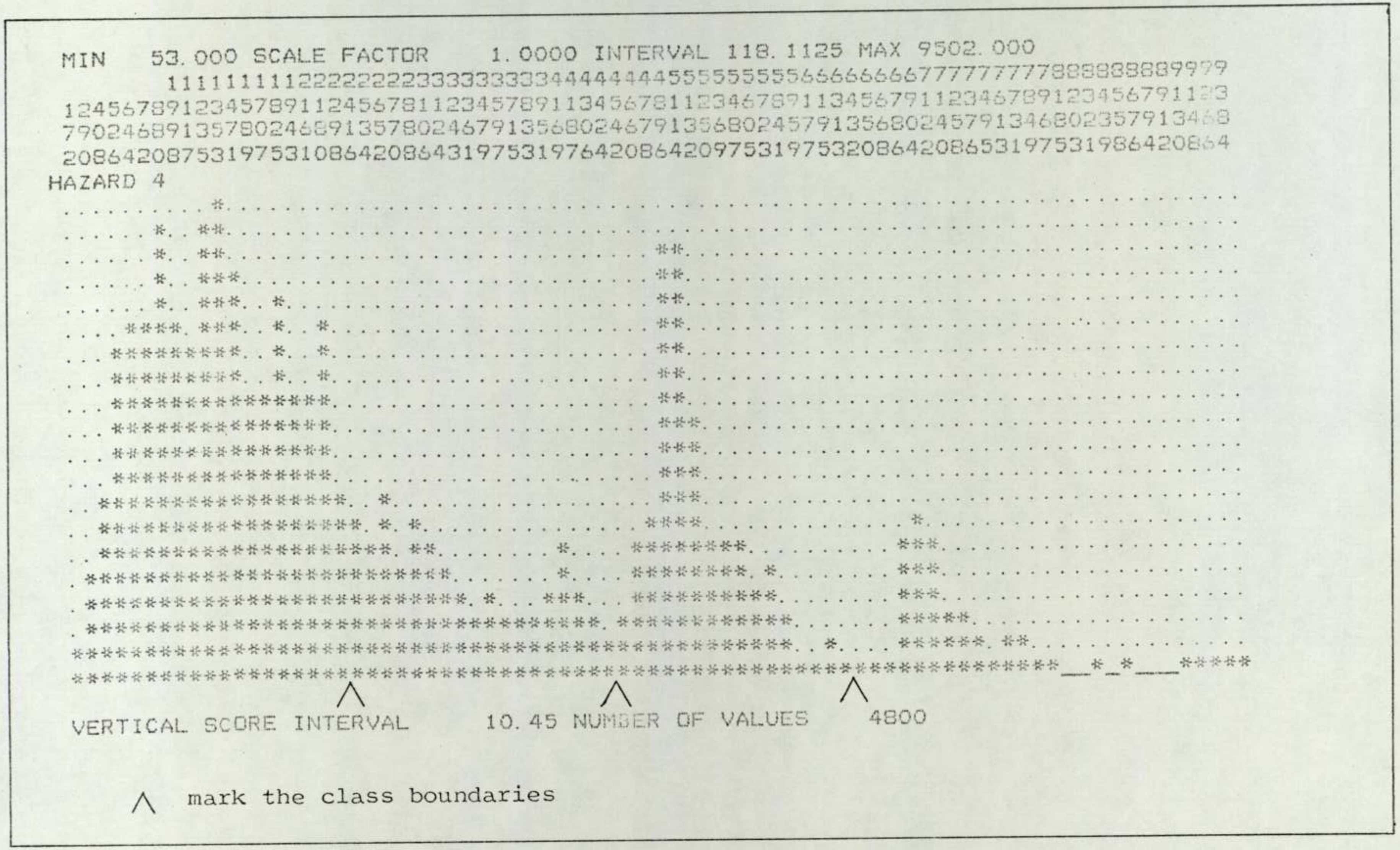
The technique of computer aided PSA is reasonably easy to implement on the map system. EXTRACTion and COMBINation to provide the required objective parameters. SCATTER to score the objectives. COMBINE to sum and weight the maps and a final SCATTER to classify the output map. The proximity, after being carried through the COMBINation, may be used to alter the intensity of the shading of the hazard colours if the DRAW output route is used.

Although each stage of this operation might seem rather empirical the result is remarkably detailed for the relatively limited effort required. It provides a good trade-off between effectiveness and effort and is well suited to the map manipulation techniques provided in the map database.

8.7 Example Two, The Map Set, Uses And Cautions

The set of maps provided as Appendix 7 is intended to be an independent collection of useful information acting as an advertisement of the map database and as an adjunct to the original site investigation files. It is made up of ten sheets. Nine are on plastic overlays for the tenth, which is a 1:25,000 Ordnance Survey map of the area covered by the maps.

Figure 8.12 Plot of the SCATTERgram for the Hazard Map



The overlay maps are of the following subjects:

8.7.1 A point distribution map

Centred symbols at the sites of segments containing the original data used in the interpolations. The symbols are sized according to the number of boreholes in the 100m square.

8.7.2 A bedrock geology map

Interpolated from the stored information, this map shows the problems associated with the Thiessen polygon interpolation. Poor data distribution tends to cause oddly shaped boundaries and the unlikely inliers and outliers are a result of rogue values.

The Birmingham fault was not considered in this processing, as an attempt to define its position from the stored geology. The result is not good, although this may indicate no more than poor reading of the I.G.S. map by the original site investigation engineers.

8.7.3 Superficial geology overlay

The derivation of this map has been described in Section 8.6.2.2. As a lithological indication it may be of more use to site engineers than the existing map, which is based on an attempted stratigraphic division of the glacial deposits of the area.

8.7.4 Shaded isopachyte map of the Glacial

The map is made up of sized symbols. The size represents the thickness and the symbol indicates the predominant grain size of the deposit. Note the thick deposits, especially infilling the proto-Tame valley.

8.7.5 Shaded isopachyte map of the Recent

Similar to the glacial map, this map uses the same symbol set and the same scale for the symbol sizes.

8.7.6 Shaded isopachyte map of the Made deposits

The thickness scale is identical to the other maps but the symbols indicate the quality of the man made deposits. The made ground is often indicated to be thick because of the localised effects of embankments but in some areas, such as tips, extensive thick deposits do occur.

8.7.7 A rockhead surface

The database model's top of the bedrock surface, including the worst case information from the overlying strata. This is a map of the level of the deepest possible occurrence of the bedrock. The contour interval is ten metres.

8.7.8 The present day surface

This is taken from the Ordnance Survey's 1:10,000 map. It is presented with a ten metre contour interval.

8.7.9 The hazard or index map

The construction of this map has been described in Section 8.6.3. The colour provides an indication of the relative difficulty of development and the size of the symbols provides an indication of the reliability of the information. This proximity can be taken as a worst case distribution map for the whole database, because of the large number of discrete parameters used in this index (or summary) map.

9 CRITIQUE, FUTURE WORK AND CONCLUSIONS

9.1 The Accuracy And Resolution Of The Map

The resolution of the information held in and produced from the map database depends upon three factors: the accuracy of the initial data, the effects of the processing it receives and the precision employed in the final display.

9.1.1 Initial data

Site investigation data are derived from several sources and are variable in their accuracy. A borehole's position in space is defined horizontally on the National Grid and vertically with respect to the Ordnance Datum (O.D.).

Coordinates estimated from the National Grid are usually said to be accurate to ± 5m. However, in a city which is constantly being re-built the reliability of the estimated grid reference can often be called into question.

Vertical accuracy may vary between a centimetre for levels near a bench mark to ten metres if map contours are used and the topographic variability is great. During the collection of information for the current project a small proportion of level errors as great as 20m were noted. Other dangers are redevelopment, which often removes - or worse, shifts the available reference points, and occasionally boreholes are recorded without an O.D. level.

The description of the soils and strata penetrated in ground investigation boreholes (their types and boundaries) is influenced by the care of the team and drilling method. Good descriptions can only be achieved when an undisturbed sample is recovered and the levels of stratal changes cannot be estimated, to better than ten centimetres, unless a continuous core is taken.

Engineering information, derived from the results recorded on site and in the laboratory, often has its precision criticised but it is to be hoped that results are generally more reliable than the 20% variability recorded by Bennett (1979) in a set of (dynamic cone) Standard Penetration values.

Translation of site investigation information into a machine readable form requires two steps. The first is an interpretation of the schematic, verbal and numeric description of a number of different investigation formats into a totally numeric form. For this step to be carried out accurately the original report must be understandable and reliable. The coding technique should either be capable of the same detail as the report or, at least, its shortcomings well known and fully understood. Finally, the observer must be neutral in judgement. The second part of the process is the precise typing of the recording sheets. In the project this has been aided by avoiding verbal implications in the codes, providing good column delineations and compressing the data fields.

Checking the information is easier once it is on the computer. The verification process has been described in detail in Chapter 6. Data justification, translation, best guess replacement of O.D. levels, test results and values which are suspect (by comparison with similar, adjacent boreholes) hopefully reduces not only coding errors but also some of the earlier misreadings of information. Further techniques, monopolising on the relative abundance of information in small areas, are discussed in Section 9.1.2.

A specific problem which deserves special mention is that of the lithological description parameter for each geological layer. A decision was made to directly code the verbal description, which changes (for the same lithology) with the age of the study. Employing these original descriptions preserves the integrity of the borehole record and allows a direct compatibility with the original report and the database map. It is also intended to provide enhanced information if a computer aided extrapolation of the map is attempted. Unlike standardised or proportional codes, use of the original lithological terms provides no barrier to conversion to other formats at a later stage because no information has been lost. The only real disadvantage in the use of specific codes is during computer interpolation; which is not, in any case, considered a very realistic tool for geological interpretation. However, even during interpolation the broad composition of the sediment is not lost. The value calculated provides a broad summary of the probable lithology predicted, as coarse/medium/fine, organic, made, soil or rock. This is considered to be adequate for most planning purposes.

9.1.2 Segmentation

The chosen structure for the database, and the concomitant capture and processing techniques, is based on an artificial limitation of the resolution of the original information. Because of the decision to limit the resolution a number of averaging reductions take place before the data reaches its final form in the datastore. Each of these processes has a particular effect, not all of which are deleterious to the end result.

When the position of a point is recorded the retrievable location depends on the precision of the coordinates stored. As Figure 9.1 shows, the collection of data on a 100m grid causes effective horizontal shifts of up to 70m in the position of point data, even allowing for the implication that all information is read as for the centre of each segment (N.G. + 50m). This translocation of point information causes smoothing and feature-shifting between a surface derived from the reduced data and the original real surface.

If the averaged samples are now considered as representative points for each 100m square in the grid another problem becomes apparent. The minimum wavelength of feature which can be resolved by a 100m grid, its Nvyquist wavelength, is 200m. From Figure 9.2 it may be seen that any smaller features of the surface are not discernable from the captured information. In fact, a more realistic minimum wavelength would be 250m because, as Figure 9.3 shows, we cannot guarantee the synchronisation of the grid and the troughs and peaks of a 200m feature.

The averaging/translocation and Nvyquist problems do, to some extent, cancel each other out. In an area of high data density each final point is an average of several values. This means that the data have been filtered and thus the recorded values are a representation of all the information in that segment, rather than a single point on the real or artificial surface through that point.

Features with a higher wavelength than the sampling wavelength are interpreted as noise on the surface. Because the sample in this case is made up from an average of several values and not a single point value, taken from the centre of the grid square, the higher frequencies are

Figure 9.1 Translocation

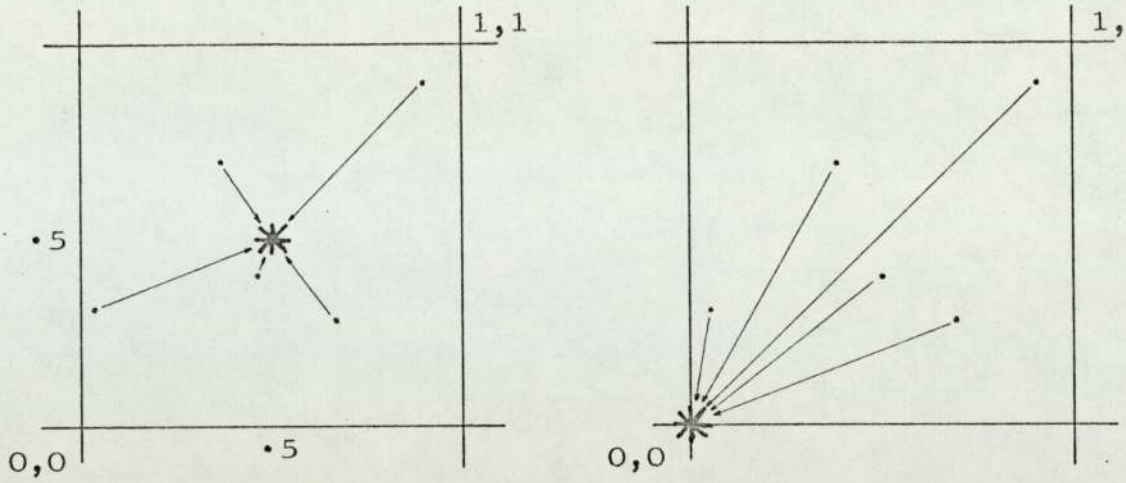


Figure 9.2 Filtering

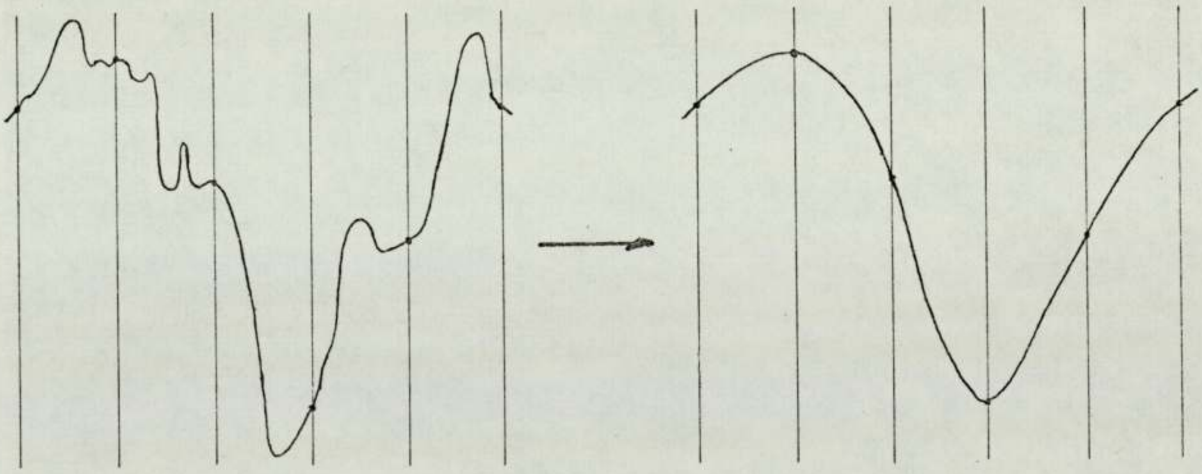
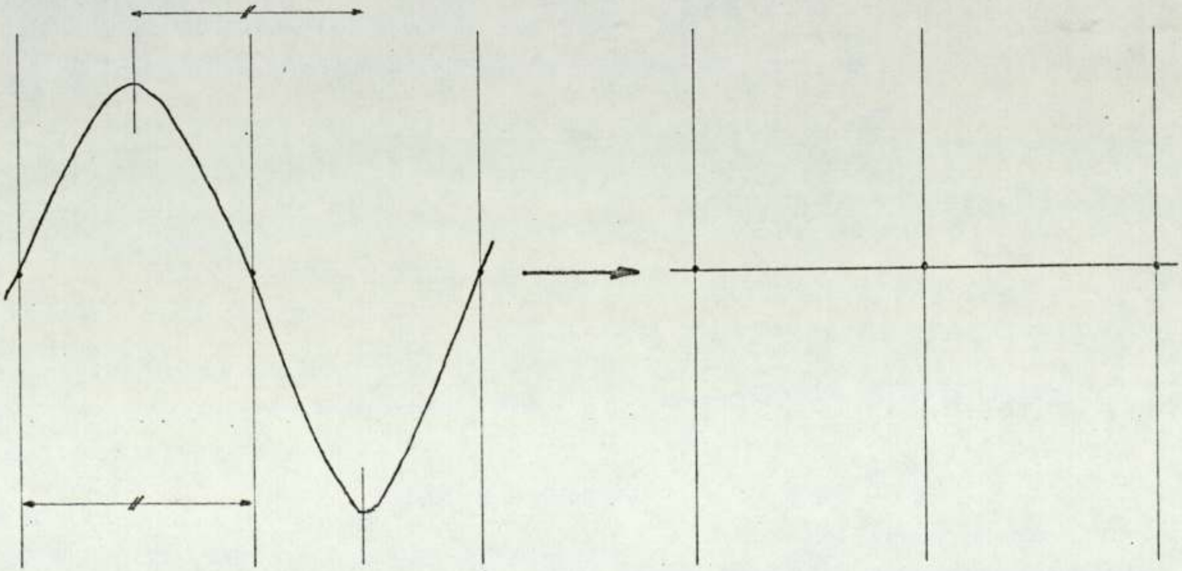


Figure 9.3 Loss of data at the Nvyquist wavelength



effectively filtered out of the data and this results in a lower noise level on the surface.

Averaging of the various types of data was designed to make the best use of the site investigation records. The treatment and some of the sources of the inevitable errors in this information have been described in Chapter 6. The data reduction techniques employed allowed the alteration of obviously incorrect values rather than the omission of untrustworthy results.

The final result of data reduction was intended to be a representative synthesis of the different ages and standards of report for each map segment. Experiments proved that a computer-aided manual technique was the most satisfactory method of achieving this result. The technique used exploits the brain's ability to quickly process visual information and extract basic patterns from complex spatial data. The results are obviously coloured by the unconscious bias inherent in both the displays used and the user. But, as a solution to an unusual problem, the human quick guess and computerised value selection offer a fast (an estimated fifty man hours for the synthesis and checking) and simple optimised technique which produces acceptable results. Another advantage of a manual method is the quality control inherent in having a poorly defined task for the computer continuously monitored by the more knowledgeable user.

When sets of layers in the various boreholes in a map segment have been defined the preparation of typical test results for the archetypal segments, from the existing information, can take place. Because of the poor precision of these data an averaging rather than a worst case selection is indicated. It is, however, desirable to keep some independent record of the worst results. These help to indicate both the reliability of the source data and/or the variability of the actual strata. A two stage process of averaging the layers and then selecting the worst averages for each broad stratigraphic age group represented generates a reduced number of records, to save space, without greatly affecting the usability of the stored map, which is in effect a worst case summary of the area.

The possibility of introducing a vertical segmentation of the whole map area, with the resultant increase in processing speed and flexibility of data retrieval, was discussed. This technique removes a lot of the realism from the information (but without greatly affecting the accuracy of the data)and introduces problems of structure and priorities, for instance: should real engineering and level parameters be treated in the same way, or, should the test results be divided up into classes and treated in the same way as the geological information. The problems of the structure and wasted storage space, for reasonable resolution, and difficulties in teaching the 'box' segment concepts of the store to users caused this next stage to be abandoned, in the current database.

9.1.3 Interpolation

One of the intentions of this project was to spread the available information by a computerised interpolation technique. These routines average either fixed numbers of points or points from fixed areas. A feature of these methods, detailed in Appendix 1, is the distortion of the final surface if the data are clustered around a few locations. Each point is given an equal weighting and the resultant surface values are skewed to the values of the closest group. Therefore a random, or even ordered, data distribution would suit this process best.

A statistical study of the data distribution from an area of the map indicates that the site investigation data are clustered and non-random. The segmentation of the information was inspired by this basic requirement. The map segment size was determined by a series of experiments to find the grid mesh which performed best in converting the existing site investigation distribution to a statistically and practically usable form. The final 100m square grid achieves a balance between statistical significance, passable resolution in practical experiments, efficient data storage and compatibility with the National Grid.

Selection of the weighting function for use in the interpolation was the subject of a similar, but more subjective set of experiments. The $\frac{1}{\text{separation}^2}$ selected produces realistic results and suits the requirements of the project better than the Shepard function (1968), which is designed for more initial than final points, or the commercial routines, which require a large number of points and are rather rigid in use.

It should be noted that this simple interpolation will only fill the calculated mesh with values between the highest and lowest provided. It will not extrapolate valleys or peaks.

Integer information may also be interpolated, to produce proximal maps. These are computer generated predictions of likely values at particular locations and although useful as such, should not be treated as anything but the most basic of interpretations.

Certain restrictions have to be placed on the computer based spreading of information over a mesh. The accuracy of the information calculated at some distance from real data points is questionable. In a geological environment it is sometimes unrealistic to allow the interpolator to use points from unspecified areas of the map to calculate new values. Facilities to place barriers on the interpolation and an indication of data proximity are provided in the system developed for the database.

The results from the experiments with sample interpolations indicate that the resolution of diagrams produced from data averaged onto a grid may be artificially enhanced, by re-interpolating onto a finer mesh grid. This line of study was not pursued, but it does promise well for the display of small areas of the map as contour or mock 3D plots. It must be remembered that this is not "getting something for nothing" and the apparently short wavelength features recovered are only a coincidence of topographic continuity and the 'realism' of the interpolation function.

9.1.4 Display

Display resolution problems are immediately apparent to the viewer. In this they would appear to be less insidious than the other limitations on resolution but they do occur and take several forms.

Contour maps are usually assumed to be continuous between the contours. The same cannot be said of an isopleth map because the boundaries, not the values, are indicated. This places a limitation on the resolution of a real value map plotted in isopleth form, even though there may be more classes than there would be contours on an isoline version of the same map. However, this limitation to the precision of an isopleth map is not always a disadvantage since it can be employed to mask noise in the information displayed by group classification of the data.

Mock three dimensional representations are usually used to display continuous surfaces. They offer advances in immediate understanding but, by definition, limit the amount of the mapped area which can be seen at any one time, due to the obscuring of some parts of the image by features 'closer' to the observer. There is also a problem with the vertical scale of such plots. A highly variable surface often obscures itself to such an extent that its underlying form is lost. Measurement of the actual (or even relative) magnitude of features is difficult, especially if a non-linear vertical scale is used.

The drawing area of a display device and its minimum line/symbol size influence the style of a presentation. Text producing devices can be used to show regular isopleth maps but the X and Y scales are rarely the same and few levels of density are available. Raster displays offer a regular X,Y grid of much smaller squares. There is still a resolution limit, however, and in programs such as SECTION (showing geological information) it is soon reached, unless care is taken and the reality of the display subtly altered. Drawing devices, such as plotters, offer greater resolution only if the pen is thin and the drawing area large. In this case the picture usually appears so faint that the bright, inaccurate raster television image is often preferred. Solutions to this conflict are only achieved by spending a great deal of money and/or time on suitable presentation and devices.

9.2 Data Storage

The current system is a practical response to a requirement for the fast, compact and simple storage and display of a multi-parameter map of a specified area, derived from defined sources. It is not a general purpose structure but the output modes and types of information are varied and therefore the store and the data must fulfill certain requirements.

9.2.1 Alterations (+-)

Additions, with their associated subtractions, of new segments and changes of information within a map segment are catered for by the system via EDITOR and SEGMENT. These allow limited revision or extensive correction of the map. Updating of the files which make up the map, to check and tidy the physical structure is available via REPLACE. This same routine also enables recreation and alteration of the files. Extensions to the filestore are possible, both by addition of single parameters and complete new sets of information. This would require some re-programming by an experienced user.

9.2.2 Ease of use

The facility with which the system may be used depends to a large extent on the usability of the computer and its operating system. The variations in speed and access times have been described in some detail. They are probably familiar to any user of a small teaching computer for large scale work.

One point which is important to note is the proportion of the total Harris 500 which the system programs and data occupy. In a compressed form (with no work files, compiled programs or backup stores) the system takes up about ½% of the total Harris filestore. In use this share increases to some 10,000 sectors, or 1% of the 300MBytes available. (NB this is still without any security copies of the files). This three MBytes may seem relatively small, but it is twenty times the standard user's allocation on the current computer.

The programs in the system have been designed to converse with the user. They offer menus of operations and prompt for responses which indicate the operator's choices. The routine housekeeping operations, such as compilation of programs and assignment of files, are carried out by small programs which operate as part of the operating system command functions. Each function is initiated by a mnemonic name and although a fair degree of flexibility is offered the primary functions are based on the usual map-derived displays of information: isoline and isopleth maps, mock 3D plots, sections and single borehole records. It is hoped that this system is not only easily usable by an individual with some experience of interactive computers but also that the function set is rich enough to prevent the experienced user from exhausting its possibilities too quickly.

The files have a simple index - pointer structure which allows a hierarchical retrieval of information. Manipulation of this structure for anything other than a spatial access mode was not designed to be easy. But, the addition of new sets of data in that spatial sense is not difficult to achieve. In this sense the database has a degree of openness. However, expansion is only available to a user who is prepared to re-program the self-contained data handling sections of the relevant programs. This difficulty occurs because the current implementation of the system is a 'first draft', intended to show the basic properties of a computer 'map' database in the time available, rather than a complete system which would take at least another year to program. For the same reason the error and excursion handling of the programs is limited to self-preservation for the datastore instead of a full safety net for the user.

9.2.3 Coding

Two forms of encoding are used for the information held in the database, namely numeric equivalent and data packing. The equivalent coding is carried out manually during the capture of the data.

During borehole coding text descriptions are looked up in tables of numeric equivalents for each geological parameter. This system is flexible and fast in use. By careful structuring a generic / specific division is made in important parameters. Permanent or temporary reassignment and the merging of several types can be achieved by patching the relevant integer codes. Extensions are simply a case of adding the new code to the lists of equivalents.

A look-up naming system is not completely compatible with the nuances of geological information which can aid the geologist to a full interpretation of the structure and history of an area. For this sophistication the proportional coding methods have benefits, despite the greater difficulties in their use. However, unlike the field slip and notebook, the information captured for the current map is not directly derived from real samples. For this study the main source of information is standard descriptions of boreholes, which are themselves gathered under less than perfect conditions. In this case very little extra loss of detail is incurred by the use of a numerical encoding technique.

The data packing takes place during the insertion of values into the datafiles. Both this and equivalent coding are well known databasing techniques which improve the storage and transfer efficiency of the information held in the computer's filestore. Standard formats are employed to aid decoding when the information is read out of the store. On fast, modern computers virtually no overheads are incurred by these extra arithmetic operations. The only safeguard required to ensure the security of the data is a check that new values are within the range permitted by the character width allocated to that parameter. If this operation is not carried out contamination of the adjacent data fields in that particular computer word could occur.

9.3 Access

9.3.1 Storage

The facility with which an intending user of the map may obtain access to the host computer, map system programs and stored information is critical in an assessment of the usability of the system. There are two possible permanent locations for the files containing the system.

ONLINE storage. Instant use may be made of the information. The files are stored on the user's disc of the computer. As has been noted the files for the system take up a considerable amount of space. Naturally, if the system is in near constant use (for instance, during the development stages of the project) this allocation can be justified. However, if enquiries are made only infrequently the space could be better employed in holding more generally useful data.

Pressure would doubtless be exerted to ensure that this soon became the case.

OFFLINE storage. This is more sociable in the computing environment. The files are copied to a machine-readable medium and the user's disc versions are erased. To re-use the system the disc, tape or paper tape which holds the information offline has to be read, by the computer, back into re-created files in the user's newly-allocated file-store. This process takes some time and there is a small but significant chance of errors occurring during the transfer.

In practice the location and security of the system should be of no concern to the user. The resurrection and maintenance of the files is the job of the computer representative in the originating department.

9.3.2 Documentation

The ability of a operator to use a system is greatly influenced by the documentation provided for him. A limited internal set of information is provided in the map database. This is intended as an aide memoir for the experienced user. Chapters 7 and 8 of this document provide more detailed information and this is, hopefully, presented without too many specialised terms. As an aid to the interpretation of the programs a set of algorithms is provided in Appendix 3. A list of all the codes and formats used makes up Appendix 2. These sections should provide sufficient information for the potential user to feel confident in commencing a session with the system. The development of a USERBASE (a group of experienced users) will no doubt extend and improve the documentation and soon provide more examples of use than Appendices 5,6 and 7, or the illustrations used in the text.

9.3.3 Hardware and software changes

Computer systems change over a period of time and long term use of the system will be influenced by a number of factors. The operating system of the computer is a large and complex program which is continuously being updated. In the short term these alterations are designed not to affect the appearance of the system to the user and thus the operation of his programs. But major streamlining of the operating system often leads to subtle changes in the interpretation of the standards of the system. When this occurs programs which depended on some peculiarity of the old system suddenly stop working for no apparent reason. This failure generally requires re-programming of macros, and possibly programs, which employ non-standard techniques.

The replacement of peripherals may also lead to difficulties because commands for the old device are inapplicable to the new machine. The only solution in this event is the re-writing of the relevant programs. The use of subroutines for peripheral control statements in programs simplifies this task and if a commercial high level library of control routines has been employed updating should be trivial.

Programming languages rarely change and if a common language such as FORTRAN is chosen, and the standards rigorously adhered to, no problems will be experienced. Unfortunately, almost every implementation of a familiar language has specific enhancements for the particular machine. These are very tempting when development time is limited, especially if the language is not perfectly suited to the requirements of the programmer. In the event of a machine change, and these are inevitable with the advances in computer technology, the extensions to the language will almost certainly alter. When this occurs the programs will have to be revised, although if the same make of machine is purchased the manufacturers may provide an interpretation program for old routines.

9.3.4 Transport

Transportation of the system creates similar problems. Moving a program or data from one machine to another within a computing facility should not present too many problems. The characteristics of both devices are known and staff with experience of such operations are available. Despite these factors the effort required should not be under estimated, because all of the problems mentioned above will have to be overcome. When transfer to another site is made, geographical separation, unfamiliarity with either the old or the new computer and lack of competence with the actual programs all have to be added to the difficulties which will be faced.

Movement of the data alone, after conversion to serial (and ASCII) form, may be accomplished with reasonable facility. Transmission by direct line from one site to another will be possible if a network is provided. In this case the original host computer and programs could have been tested at the remote site, via the network, to ensure suitability. Other means of data transfer depend on the transfer of the information to a storage medium which can be read at both sites.

Transfer of programs is more difficult and in general the routines produced from small projects are not transportable, in the coded state. It is almost always more efficient to re-write programs from the original algorithms than waste time trying to remove system-specific bugs from machine readable copies.

9.4 External Users

9.4.1 Advertisement and costing

The current computer-stored map is a useful tool for the planning of site investigations and choosing sites for development in the Birmingham city area. It was designed for use by specific groups of individuals who obviously work outside the academic environment. Advertisement of and access to the system must therefore be established on a professional basis. The first stage in advertising the 'map' is the circulation of the maps in Appendix 7 to the parties approached for information or aid during the project. This distribution will announce the existence of the map but the development of a userbase is probably a matter of personal approaches to potentially interested parties.

Table 9.1 shows the costs of the various system operations, estimated at the current charging rates for external users of the Harris 500 computer. This indicates, all too clearly, that a charge will have to be made for the service provided.

In the early stages of commercial use, however, the service charges could be dropped as an advertisement. This would test the usability of the system and might solicit payment in the form of new information to enhance the map. If the system proves reliable and popular enough a scale of charges can be introduced. This will, hopefully, not be resented when the client is already aware that the information will be useful and, even more importantly, reliable. The cost of the free trial period and the map set* cannot be

---

* circa £70 per set. It should be noted that even as long ago as 1973 SYMAP plots were costed at $4 each.

Table 9.1 Costs of the Database Programs

PROGRAM NAME		TIME ec CPU	COST £	NOTES
EXTRACT		50-80	·50-·80	dump takes 7 minutes
E/PXTPOL		180-540	1·80-5·40	depends on the number of points
REPLACE		180	1·8	a restore takes 7 minutes
COMBINE		10+5	·10 ·05	first map is 10p every additional map required is 5p
SCATTER		6·5	·07	includes a paper copy
CHORO		8·7	·11	includes a paper copy
DRAW	compile	21	·21	only required once per session
	3 D	16	·16	user provides paper for the small plotter, large plot extra (full screen)
	line	25	·25	
	pleth	50	·50	
SECTION		·75	·30	time is for 1 well on paper, cost is for a 40 well paper copy
SEGMENT		5·5	·04	time for one well, costs decrease with number to circa 2p ea.
STARTUP		43	·43	operation is essential only for a display session
PSOUT		240	2·50	1:25,000 underlays are circa 10p each
EDITOR	compile	20	·20	only required once per session
	per well	·5	½p	

retrieved but should be written off, along with the estimated £13,000 of development computer time and the additional non-computer costs of the map database.

9.4.2 Enquiry sessions

External users will not access the system frequently and cannot, in any case, be expected to spend a large amount of time acquainting themselves with the necessary commands and functions. Most complex devices, such as microprobes or literature search systems, have a technician to provide the routine expertise while the client gives instructions. This would seem to be a suitable solution to both the costing and experience problems for the database. A departmental user could be trained to use the system quickly and efficiently for external users. The machine time saved would almost certainly cover the extra cost of the assistant and charging for the resources used in the session would be simplified. The client would benefit because he would be sure that he had achieved the best results for his investment.

9.5 Extensions To The System

The experimental nature of the present implementation of the map database has been mentioned above. During the construction of the system certain features were planned but had to be omitted due to the shortage of development time. A number of these, reasonably simple, enhancements to the system are listed here. Because of the segmented, or subroutine, approach employed in the creation of the programs and the simple hierarchy of the data files few difficulties should be experienced in the implementation of these programs.

9.5.1 Individual programs

Appendix 3 contains a description and list of recommended enhancements for each program. Some examples are as follows:

The SECTION vertical cross section plotter can be altered to use a general purpose set of colours, specified by a VLT, and textures. This would be in addition to the geological standards currently employed. These new shades and symbols may then be used to electively represent any set of parameters held in the database as a cross section. For instance, moisture content in blue to red and qualitative strength as shading, to highlight incompetent strata at depth for an extended site across a river valley.

SEGMENT can be enhanced by the adoption of features from the data reduction programs. Averaging of old and new information, rather then the replacement of records, and a less rigorous entry format would be useful improvements if the map database is to be maintained. If expansion is contemplated a batch insertion mode would prove very useful, drawing data either from cards or the data logger described in Chapter 6.

The interpolation programs EXTPOL and PXTPOL provide proximity surfaces to indicate the local reliability of the calculated map. An additional overall index of the quality of the interpolated surface would often be more convenient. Such an index could be calculated by the program from: the number of initial points, the number of points included in each average and the mean separation of the initial points. This value might be added to the calculated map file and preserved, or recalculated, during subsequent processing for an indication of the quality of the final map.

9.5.2 Total system

The programs which make up the functions of the database are, at present, largely independent. Certain broad macro-scale operations would aid integration of the separate operations into one integrated system.

An example is the addition to all the programs dealing with map files the ability to extract, process or display limited areas of the stored map. This alteration requires a limitation of the looping of the map handling routines. The means by which the location and size of the map are transferred could simply be the map file itself. However, if a central register file were used instead this would open the way for a general construction of 'softer' controlling macros and better fail-safing of system programs. This central register could be designed to hold file assignments, parameters and the critical indices for long-running programs. In addition to these simpler uses a central register file might be programmed, by a suitable language interpreter, to hold a number of system function steps. This would allow the 'macro scale' programming described in Chapter 8 to be employed directly, thus eliminating a gap in the conceptual description and the actual use of the database system.

9.5.3 Statistics and filters

A useful extension to the handling of the geotechnical information in the database would be a statistics package. The ability to take parameters from specified areas of the map mentioned above, in conjunction with the interpolation barriers would make this package more flexible. One useful feature might be the ability to display variables against each other, the program would operate in the manner of the

SCATTER plotter. These plots can aid understanding of the nature of the deposits, perhaps for example, providing enhancements to the lithostratigraphical subdivision of the map, after Horton (1975). However, the spatial nature of the data would naturally be lost in this process.

Analysis of the spatial statistics of the various data distributions will aid the proper use of the interpolation programs. Spatial data filters, based either on harmonic analysis or superimposed matrix techniques, are generally available for analysis of the structures in an area. While the study area is too small for this sophistication a simple patch smoothing function could be employed to advantage in 'cleaning up' maps prior to plotting. In this process a smoothing array is repeatedly recalculated from, and superimposed on, the map to eliminate (probably) spurious spikes in the source map array. With carefully designed superimposition rules the smoothing of real edges can be minimised.

9.5.4 Vector overlays

A feature of the handling of paper maps is the ability to introduce new data on transparent overlays. The DRAW program can to some extent provide this facility but only for isoline and isopleth maps. The addition of maps of other types of line drawn spatial information, such as housing or transport routes, is not possible; unless a copy of the DRAW map is made on film for registration with an existing e.g. Ordnance Survey sheet. CHORO has no possibility of superimposition over its paper plots which have different X and Y scales.

If a common vector plotting code is employed, digitised overlay maps of the required features can be held in a common store and plotted over either CHORO or DRAW maps. Since only the Sigma graphics unit will provide the colour necessary for discrimination between the overlay and the original map the plotting code can be device-dependent and therefore fast and simple.

Encoding of vector overlay maps, for storage in files associated with the main datastore, may only be practically accomplished with a digitiser. A television camera might be employed to capture data more economically but the current graphics device is not capable of superimposing two raster images and no 'frame grabbing' raster digitisers are available.

9.5.5 Refresh graphics

The possibility of coupling a refreshed vector plotting graphics output terminal to the map database has already been mentioned. GINO-F is very suitable for driving this type of device, thus providing the implicit ability to make hard copies of any screen image.

An interactive enhancement program for the stored model of the mapped area has been proposed. This would operate in a similar fashion to the French VERCOURS program, created by Buisson et al (1979). A small area of the stored data is transferred to a work file with the same structure as the main datastore. Two display modes are provided. The first allows the three dimensional surfaces which make up the current model to be shown and moved independently. This display is intended to aid the user in visualising the structural relationships of the strata. A second mode of display (directly derived from the existing SECTION program) provides the user with the ability to view selected cross sections through the area and to alter them with a light pen. The boundaries can be re-drawn or dragged with the pen. The existing lithological information might be used to fill sediment lenses in the model or new information may be inserted by the user. A simultaneous coloured illustration of the section on a raster VDU could be provided. This would be re-drawn at set intervals to show any new alterations.

Successive areas from the map may be withdrawn, enhanced and then replaced in the original, or written to a new, store in order to build up an improved map model. Naturally, the provision of interpolated geotechnical data in an initial simple structural model is a necessary aid to this manual interpretation.

Other uses of the refresh display in the map system involve simulation. Cartoons of the sedimentary history of the area can be animated. Architectural and Civil Engineering display programs (such as Abel and Kovack's computerised model of Chicago) may, for the first time, be extended to include the soil and subsurface features of a site. And, of course, the map database can not only provide images of, but also predictions on the behaviour of the soils shown.

9.5.6 Extra files

The connection of files of new information into the datastore structure is a simple matter. A REPLACE operation will introduce the necessary pointers into the master file and transfer a serial file of new data into a direct access datastore file. Typical extensions might be: the addition of contractor's information for each borehole, provision of a sample log or greater archive detail. All these parameters are point data which may be collected on the 100m grid of the present map.

Vector information has been illustrated above with the resultant provision of overlays of, for instance, Ordnance Survey maps. Full exploitation of such data requires an ability to access the information by the same Segment, Section and Map (of any area) functions as the basic raster data. A simple structure for such an extension to the database is as follows:

A register file, accessed from the master file of the database, holds the coordinates within which each stored line can be contained (or BOXED), their start coordinates and type (house, river, sewer etc.). A set of subordinate files contains the digital descriptions of each line or feature. These are held in a code such as Freeman's chain code, Freeman and Peroni (1979). The image to be retrieved could thus be specified as to the area, by a boxing test for included lines, and the type of information to be displayed.

9.6 Re-Use Of The System

The structure developed for the map database may be altered in a number of ways. This makes the interpretation programs now in existence, or created in the future, usable with other sets of data than are provided at the present.

9.6.1 Change of scale

If the map database is employed to hold other types of geological data different grid sizes will be required. This scaling carries with it an implicit alteration in the vertical resolution of the database. Various sizes of grid mesh and vertical scale suggest themselves as ideal for certain types of data. Some examples are as follows.

A one kilometre grid is useful for county-sized maps.

On this scale the store will be useful for the preparation of smaller scale planning maps. The inclusion of all Phanerozoic strata will enable the system to be employed for broad scale structural analysis. Ten kilometre grids would be suitable for National studies, following the guidelines of Rockaway (1976) and Matula (1978). Extension to a 100km grid will reduce the usefulness of the vertical dimension of the map but the system could prove useful for continental registers or tectonic analysis. The 'height' could possibly be replaced by some other dimension, such as age of strata.

A decreased grid size opens up the possibilities of representing steeply dipping strata. A 10m grid would be more useful for small structural models and a 1m mesh allows the database to function as a site log. Vastly reducing the mesh to a millimetre scale reveals the intriguing possibility of recording data from hand specimens or thin sections on the database.

9.6.2 Change of parameters

The basic structure of the database is capable of recording and displaying for analysis any spatially related sets of data. Some examples are as follows:

LITHOLOGICAL information. An increase in the mineralogical and chemical data recorded will enable sedimentary and metamorphic facies interpretation.

ENGINEERING parameters. Grading curves and consolidation tests with improved compression test results could allow more sophisticated pre-investigation study of potential sites.

HYDROGEOLOGICAL information. A file of dated water levels and permeabilities for specified strata would not only facilitate the cheap production of water maps but also be of great use in the interpretation of geological structure.

STRUCTURAL data. Storage of foliations and orientations is essential in metamorphic mapping or structural analysis.

GEOPHYSICAL data. This can be used to enhance the more usual site investigation data and is especially useful in urban areas.

REMOTELY SENSED data. These are basically two dimensional but multi-spectral information might be held efficiently in the layer hierarchy. This would allow several sets of information, including ground truth, to be registered against one another. The processing and display functions are well suited to the raster format of remotely sensed data.

9.6.3 Change in dataset capture

The collection of information for the map could be greatly improved. A manual capture method is slow for several reasons. One factor is the location of the information. The establishment of a National Registry of ground investigation reports would allow the coding to take place from copies retained by the coder. This would eliminate much of the travelling and administrative difficulties associated with data collection. An added advantage would be the implicit retention of a record of the source data. This will facilitate both the checking of the information and fast access to the original data by a subsequent user. A file of microfiche numbers might be added to the datastore files to directly index this information.

Another improvement in coding efficiency would be achieved by the direct entry of information into the computer system. This could either be at a low level, via a simple prompting and storage program, or directly into the datastore via an enhanced SEGMENT program.

The use of unskilled individuals to code up samples has been studied by Griffiths and Rosenfeld (1954), Sokal and Rohlf (1970) and Chadwick (1975,1976). These authorities report that the description of specimens is the limiting factor in the reliability of the results. Because translation to a standard verbal description has already been carried out, by the authors of the site investigation reports, the use of unskilled coders is expected to be an effective step in the improvement of the efficiency of the data capture.

Translation from existing computer databanks is a possible source of pre-coded information. This route was used in the current project and the cautions on data acquisition listed in Chapter 6 should be noted before a decision to employ machine readable data sources is made.

There is a practical limit to the size of the datastore which may be conveniently handled by the Harris computer and the database system. Maps of new areas, probably on different scales, and possibly even extensions adjacent to the existing area will be more efficiently stored in separate sets of datafiles. The interpretation programs might then be connected to whichever map is required at the time. The other maps can be held online or offline in either single, serial (dump) files or sets of direct access files.

9.6.4 Changing the field of application

The present database is geotechnical. The structure and programs are suited to the storage and retrieval of many forms of spatially-related data. This section provides some examples of the use of this spatial information system in other scientific fields.

Mapping of tectonic features, with a large grid mesh, requires field and literature surveys to provide the necessary information for the datastore. Stratigraphic divisions might be based on, for example, metamorphic characteristics.

Small scale analysis could be accomplished by the encoding of field slips, cliff sections etc. from existing or new (electronically logged) investigations. The information structure might be employed for simple retrieval or to supply data for advanced fold regime analysis and simulations.

Geochemical and, after some processing, geophysical data are compatible with the map structure. Several scales of maps of these data are useful and, with the machine readable output produced by modern survey and analysis tools, the capture of such data could be very fast indeed.

In the field of Remote Sensing an "information explosion" is taking place and analysis software is lagging behind the data available. The database has a very suitable format, and already offers analysis tools for, these 'raster' data.

A more exotic use, on a much smaller scale, could be logging of data for the electron microscope or the microprobe. These devices produce spatially-related machine readable output which can be treated in a similar fashion to other remotely sensed data.

A very practical use for the map database is Archaeological site logging, on a ten or one hundred centimetre scale. In this application the source of the data, the site, is destroyed during investigation. The database could be totally filled with original data (and would be the most 'real' remnant after the study had finished). Because a 100% three dimensional coverage is achieved no interpolation is necessary. The representation is very accurate and a fine mesh is justified.

9.7 Mounting The Database On A Microcomputer

The possibility of making use of the new technology of the microprocessor-based computer in this project was mentioned in Chapters 3 and 6. This section presents the advantages of this approach and some of the possibilities and limitations of the state of the art in personal information processing.

9.7.1 Advantages

Mainframe computers are expensive to purchase and run. If the current database system had been produced outside the academic environment the estimated development and running costs would almost certainly have prevented its completion, or use. A microcomputer costs no more than its minute current consumption to run and the hardware is several factors of ten cheaper to purchase.

Access to a microcomputer is usually instant, no sessions need to be booked because the device is cheap enough to be dedicated to one function. The output does not have to queue and therefore may be printed or plotted instantly. Incidentally, fewer hard copies are needed since the device is permanently available to re-plot the more understandable original screen image.

A microcomputer may be made more user-friendly than a mainframe because the monitor, or operating system, can be designed specifically for the required task. Integration of the monitor with the programs allows the whole machine to permanently function as a single, dedicated, unit.

Another benefit of the microcomputer is its small size. The whole system may be used in the field to log results directly from the measuring instrument or borehole. As an example, Sowerbutts and Mason (1981) at Manchester have developed a system for the logging and analysis of small-site geophysical information. This system is employed in the investigation of development sites for buried man-made structures. Surveys may be studied on site to ensure the reliability of the results. This leads to improved final results and eliminates the frustrating data reduction time lag.

9.7.2 Advances

The present generation of microcomputers are slow and small with limited filestore with respect to not only mainframes but also minicomputers (such as the Harris) but advances are taking place very rapidly. Typical operational parameters are given in Table 9.2. From this it is obvious that, with the current advances in microcomputer technology, there will soon be little benefit to be gained in any basic parameter by using any other small computer than a micro for specific projects. In fact the convenience of the filestore management of the microcomputer may already make this device more popular than a shared system for some users.

9.7.3 Acquisition

There are two possible implementations of the current map database on the present generation of microcomputers. An identical system might be created, with all of the facilities, on either an expensive graphics microcomputer or a micro driving a raster graphics terminal. Alternatively, some constraints on the resolution of the display and the

Table 9.2 Typical machine parameters

MACHINE TYPE	word size in bits	clock speed in MHz	core size in Kbits	core size in Kwords	typical filestore in KBytes	typical filestore in Kwords
mini computer (Harris)	32	45	1,920^{+}	60	300,000^{+}	75,000
typical maximum user's allocation			256	8	2,880^{+}	720
current micro computer	8	1·8	16/64*	2/8	250	250
new generation micro computer	16	4/12	256	16	2,000	1,000

*usually supplied with 16K of main memory, expandable to 64K $^{+}$fixed maximum storage speed of operation could be accepted and a budget device employed. The advantage of using a cheap machine is the interesting possibility of selling a dedicated piece of hardware as a 'map'. This device would be capable of a far more flexible display than any set of similarly-priced paper copies.

Some of the obvious limitations of the micro for this use can in fact be overcome with careful programming. The effective speed of the machine may be greatly increased by using faster programming languages, such as the Assembler code of the device (which can often be compiled) and optimising all functions. Storage is actually easier on a micro because discs can be easily exchanged thereby eliminating the on/offline storage problem. If a well written disc operating system is available several 2MByte drives may be used in concert, offering more personal storage to the individual than is possible on a shared minicomputer.

The (raster) graphics are of limited resolution, but against this they are more flexible. This is because the image is stored in the core, or main memory, of the machine. Thus there is no transmission delay - the major factor in limiting the speed of computer terminal response. Advanced graphics programs are slowly being made available for microcomputers. For instance, Sowerbutts (pers. comm.) has developed a Ginosurf-like package for micro use. This could replace DRAW, the only library-dependent program, if the map database system were transferred to a microcomputer.

9.8 Conclusions

9.8.1 Fundamentals

A geotechnical map is made up from geological and engineering descriptions of all the strata in an area. The display of these many parameters has proved difficult for conventional map-making techniques. If a map is defined as: any spatially related set of data, it becomes possible to dispense with the concept of a 'map' as a physical entity. A convenient tool for the storage and display of large amounts of information is the digital computer. Therefore the basis of this project is the construction of a computer database which will function as a flexible geotechnical map.

The research thus divides itself into two linked tasks. Firstly, the production of a map from the existing information sources using a computer database. Secondly, the database and the current dataset processing are the 'data' for a study of the improvements of the concepts and realism of the database itself: i.e. a geological study of the uses and relevance of a computerised spatial information system for urban engineering geology.

9.8.2 Concepts

A geotechnical map is considered as a database. For this process to be successful the following points have to be born in mind:

The resolution of the information must be analysed to ascertain the most accurate and efficient storage techniques.

The spatial nature of the information is to be preserved either implicitly, by the structure of the storage, or explicitly, by recording of locations with respect to some datum.

The display of the information must be at least as concise and simple as in conventional maps of engineering geology.

In this project geological information is interpolated by a computer. A number of conclusions have been made about this process. Firstly, the distribution of the original data must be suitable for the interpolation routine. Secondly, the weighting function employed has to produce realistic results with the concentration of data present. Finally, continuous variables produce more realistic results than ordinal parameters.

The original information is reduced by computer-aided visual processing. The reduction is carried out to decrease spatial and inherent errors in the site investigation data employed. Computer-aided techniques were used because manual reduction of data to the database format would be too time consuming and fully programmed techniques lack the necessary discrimination. A visual display with carefully selected parameter and spatial definition allows the superior processing skills of the human mind to quickly synthesise the 'best' results whilst the computer automates the routine tasks.

The results of this process show that data can be successfully filtered to provide information of the greatest usability. Because the logical manipulation of existing data may disclose facts which were not evident from more simplistic study, techniques such as macro scale programming and potential surface analysis were used to combine discrete sets of data to produce documents which are more reliable and understandable than the original information.

9.8.3 Benefits from the data project

The original site investigation reports contain precisely located, detailed information which may not be complete and is of variable quality. A superstructure to these amorphous files has been established. The new map database contains original and interpolated syntheses (or archetypes) of all this information in one hundred metre square segments of the mapped area. As an index to this information a derived 'hazard' map provides an indication of the values and validity of the map database.

This hazard map is an interpretation of the geological, chemical, topographic and engineering features of the mapped area. It is designed to give an indication of the difficulty of development of an area, as presented by the rock, soil and man-made deposits in each 100m square. The map also indicates the absolute availibility of full information in both the database map and the original site investigation reports.

The database contains many hundreds of borehole records which reach the Triassic bedrock in the central Birmingham area. After several attempts, a reasonably realistic rockhead relief map, figure 8.10, was prepared. This map must be treated with some caution because the interpolation program employed can only interpolate, not extrapolate from, the level data provided. Therefore the final surface is not a true contemporary geomorphic surface but a sum of the late stage Glacial, Post-Glacial and Man-Made excavations into the Glacial topographic surface.

Subsequent users may be able to improve on this representation because the data used are still available. The ultimate benefit from the test data study is the creation and existence of a multiparametric, geotechnical map of central Birmingham in a database designed as a user-friendly system which is both simple to use and easy to extend.

9.8.4 Benefits from the database project

Research and development of graphics techniques have formed a large part of the project. The raster graphics output controller is a relatively new device in this particular field of computerised information display. This machine's ability to present solid areas of colour and texture on a standard colour television screen is well suited to the display of geological information.

Raster techniques developed in the project include the modelling of lithological colours and textures on raster devices and the employment of identical plotting methods on a television screen, a matrix plotter and the standard line printer.

The ability of the digital computer to store and process large amounts of information makes it the ideal data verifier. In this project the initial data were checked and in some cases corrected or enhanced automatically. This process allowed an accurate check to be kept on the frequency and type of error thus improving the reliability of the estimate of the accuracy of the final results.

A notable feature of automated data processing, including the later interpolation of the data, is the reproducible nature of the treatment given to the information. This reproducibility is especially useful when the data are to be employed by more than one individual. This is because the uniformity of the original subjective computer based processing will effectively render this process objective to a later observer, or user, of the results.

In the context of the interpolation of site investigation data no great advantages were seen in the use of sophisticated mathematical curve-fitting techniques. Simple, weighted averaging between existing datapoints is considered to be a safer and more predictable method of data spreading. The use of an experimentally determined interpolation function, matched to the available data, with positive interpolation barriers was found to provide the optimum balance between truth and full data coverage. It is strongly recommended that any enhancements to these data spreading techniques lie in the development of interactive, rather than fully programmed, extrapolation.

One of the major features of this project is the creation of a set of program algorithms for the administration of a spatial database. These routines are at present implemented in Harris enhanced FORTRAN 66 and generally produce graphical output readable by a Sigma T5680 graphics terminal. No suitable general purpose raster graphics control libraries yet exist. But it is recommended that any future implementations of the routines are written in PASCAL which is well suited to the program type and structure.

In the mid-seventies a number of projects to establish geotechnical databanks were carried out. The general conclusions seemed to be that computer stores were too expensive and inconvenient for everyday use and many of these early systems were abandoned. However, computer technology has changed very rapidly. The size and cost of both processors and storage has decreased markedly in the last five years. This project has reviewed the state of the art.

The arguments of long turnaround time and inconvenience against the large mainframe still stand. Even the modern minicomputer, despite its high speed and interactive facilities, is still too expensive in use and unless some very real hardware benefits are offered its employment can not be justified.

The mass produced microcomputer, however, offers adequate processing power and sufficient storage facilities for almost any application. In addition the low price of these devices negates all the costing arguments levelled against the previous hosts of geotechnical databanks. It is expected that interest will revive in geotechnical databases and it is to be hoped that this work will form the basis of at least one such study.

APPENDICES

a1 INTERPOLATION EXPERIMENTS

a1.1 Determination Of The Optimum Map Segment Size

a1.1.1 Introduction

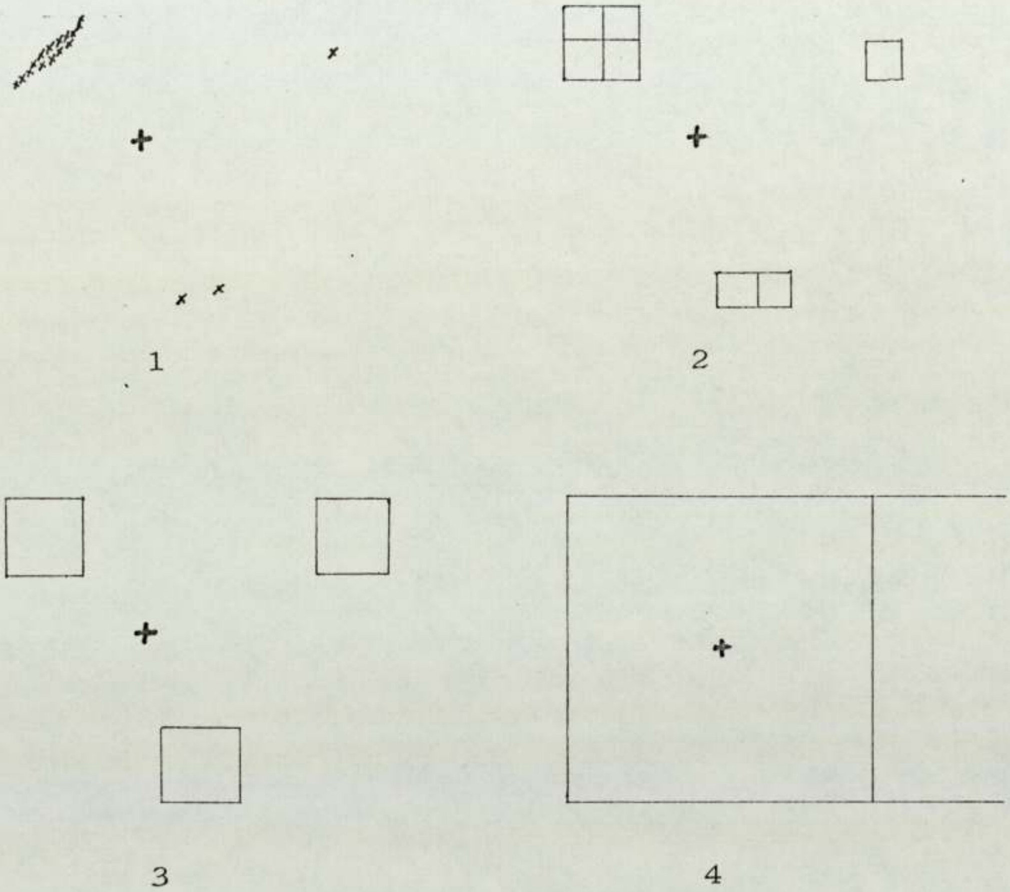
It has been proposed that, in addition to: reduction in storage space, reduction of errors and convenience of handling, the averaging of the original information on to a regular grid will improve the operation of computer based interpolation routines. These programs require a dense, random set of initial points for good operation. If a dense point set cannot be achieved then good results may be obtained if the data are randomly distributed.

Subjective and objective experiments are required to determine the optimum pixel (or map segment) size for the distribution of information available in the mapped area. This size will set the limit of resolution of the map.

The averaging of information into square pixels is intended to reduce the clustering inherent in the collection of data for discrete sites. Figure a1.1 shows the effect of segmentation on a clustered set of data. For cases 1 and 2 the interpolation of the marked point skews the calculated surface to the mean third dimension (or Z) value of the groups of points. This is because no consideration of the geographical location of the data is made in simple interpolation routines. Case 3 shows a distribution which would not skew the surface and in case 4 the resolution of the initial data has been lost.

The data used in this set of experiments are derived from Varey's databank (1977). The areas selected for the experiments are generally in the areas of better-than-average coverage and this fact should be born in mind when assessing the results with regard to the estimates of resolution.

Figure a1.1 The Effect of Segmentation on Clustered Data



a1.1.2 Experimental techniques

ONE Subjective examination of a set of maps generated from an artificially created surface at pixel sizes of 25,50,100 and 200m (square).

TWO Quadrat analysis. The initial point distribution is quantised into pixels of 25,50,100 and 200m with the original data acting as a 10m pixel size. The resultant 'data points' are plotted for each grid and the point distribution is tested for uniformity by Quadrat analysis for thirteen sub-areas.

THREE Nearest Neighbour analysis. The data distributions calculated for the Quadrat analysis were processed to find the average separation of the points in three subareas. This information is then used to test whether the distribution of the data is random or clustered.

FOUR Mathematical comparison of interpolated maps via a program (MPL). Reduced sets of information on the O.D. level of the bedrock from an area of the available data with good distribution. These are used to calculate a bedrock surface for each of the four pixel sizes: 25,50,100,200m. The resultant surfaces are compared with a surface calculated from the original data. This comparison can be presented in a number of ways: difference, percentage difference, square of the difference. A value for the root mean square (rms) difference is also calculated.

Tests are made not only by progressively reducing the number of datapoints over the whole map but also by confining the supplied information to strips of the map. This latter test is to determine the effects of no information whatsoever on the interpolation from various grid meshes.

a1.1.3 Results

ONE Subjective analysis. This revealed a predictable smoothing of features with increasing pixel size due to the effective reduction in datapoints during the averaging. An effect named translocation was noted. This is a shifting of features due to the interpolation from a newly-created 'average' point location in the centre of each pixel. Another effect which occasionally occurs, named the shared point effect, causes a 'sharpening' of the surface. This arises from the inclusion of the value for a point into two adjacent pixels because it lies on the boundary between them. This effect is more severe than might be expected since, due

Figure a1.2 Results of the Quadrat Analysis

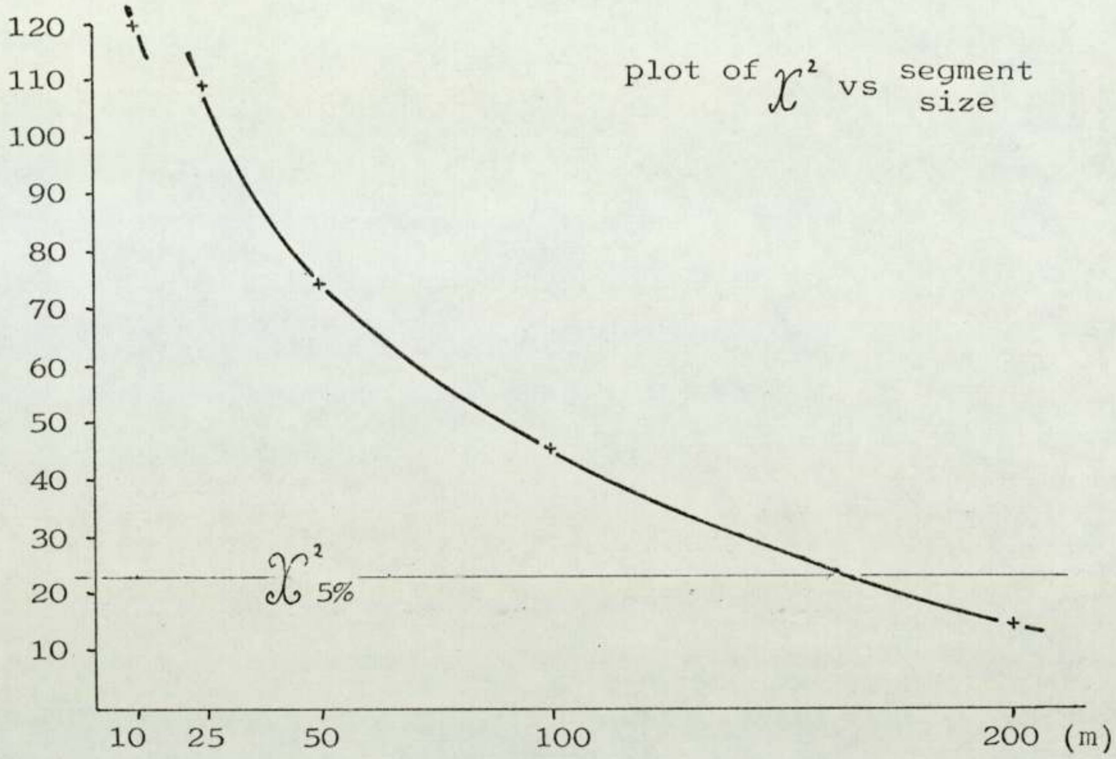


Table a1.1 Nearest Neighbour Analysis Results

segment size (m)	sample area	$\frac{\text{actual}}{\text{expect.}}$ sepn.	distribution characteristic
10 orig	A	0·58	clustered
10	B	0·60	clustered
10	C	0·63	clustered
25	A	0·60	clustered
25	B	0·63	clustered
25	C	0·68	less clustered
50	A	0·70	less clustered
50	B	0·73	less clustered
50	C	0·81	near random
100	A	0·94	near random
100	B	0·96	random
100	C	1·11	random
200	A	1·28	near random
200	B	1·29	near random
200	C	1·47	near ordered

to the limited spatial resolution of the original data, the 'border' of each pixel is 10m wide.

TWO Quadrat analysis. Figure a1.2 presents the results of this test. Only the 200m pixel size comes within the limit which indicates a 5% significant uniformity. The break in the graph for the untreated original data is due to the absence of the shared point and translocation effects mentioned above.

THREE Nearest Neigbour analysis. Table a1.1 presents the results of this test. The ratio of observed to expected mean point separation for typical distributions score as follows:

0	all points coincident
0 - 0·5	clustered distribution
1.0	random distribution
2·15	maximised or ordered distribution

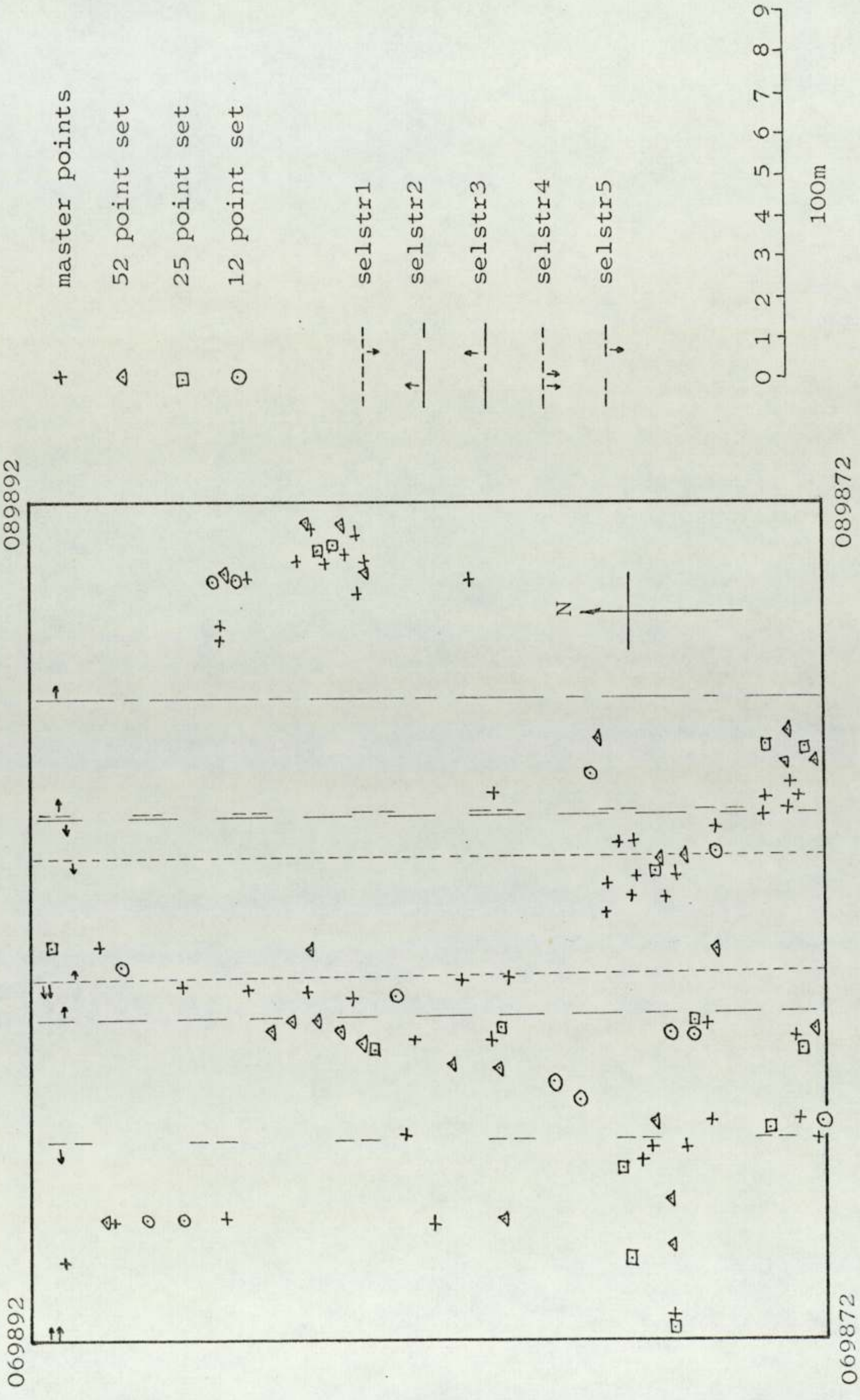
For the tested pixels and data the 100m size gave distributions nearest to random for all three subareas. These subareas had only slightly different initial distributions thus no prediction of the effect of greatly reduced data concentrations may be made.

FOUR Calculated comparison. The data distribution is shown in figure a1.3. Table a1.2 and figure a1.4 show the results for the tests on reducing the number of datapoints. Table a1.3 gives results for the set of tests in which the sampled area is reduced. After the first test on 101 datapoints the impractical 25m pixel size was replaced by an unaveraged control surface plotted directly from the reduced dataset.

a1.1.4 Experimental conclusions

Figure a1.3 shows the original dataset of 101 points. This was the best available but, unfortunately, it displays most of the characteristics of site investigation data.

Figure a1.3 Data Distribution for the Interpolation Experiment



The datapoints are clustered because of the simple fact that site investigations are only for small areas. Single points are often distinctly 'odd' as indeed are some whole (single investigation) clusters. In the case of the bedrock level the causes might be inaccurate measurement, poor levelling or misidentification of the lithological interface. The distribution of points and clusters is poor as a whole, with very large areas completely devoid of information.

The interpolation process employed is simple for reasons of efficiency and averages distance weighted 'Z' values for the six points nearest to each grid node in turn. The results are written to a two dimensional array, or digital terrain model (DTM). According to Davis (1973) six is the optimum INTERPOLATION FACTOR. The results for the smaller datasets and coarser mesh sizes tend to contradict this proposal for two separate reasons.

Firstly the small datasets have a large mean separation between the datapoints, thus averaging with six points in a sparse set leads to inordinate distances over which points supposedly affect each other. However, reducing the interpolation factor leads to discontinuities in the final surface, because of the interpolation noticeably switching between one point and another.

Secondly the coarser mesh sizes pre-average a number of points over a measurable area - similar to the Kriging process, with fixed, square fields. This pre-averaging means that the order of influence of each pixel is greater than that for each point in the original dataset. Therefore, taking six points is equivalent to averaging a great many more and, implicitly, several will have the wrong distance weightings.

Table a1.2 Results for the Reducing Points Experiment

interpolation factor = 6	measure	25m	50m	100m	200m
101 original points	points	131	119	79	42
	rms %	2·2	1·9	2·0	3·8
25km^{-2} point density	peak %	12	11	8	-15

	measure	original	50m	100m	200m
52 original points	points	52	64	44	31
	rms %	3·6	4·4	3·8	5·1
13km^{-2} point density	peak %	±12	20	17	-17

	measure	original	50m	100m	200m
25 original points	points	25	33	25	21
	rms %	9·7	9·8	10·0	9·8
6km^{-2} point density	peak %	-45	-45	-45	-45

	measure	original	50m	100m	200m
12 original points	points	12	15	13	10
	rms %	11·1	11·2	11·4	11·6
3km^{-2} point density	peak %	-40	-40	-40	-40

interpolation factor = 3	measure	25m	50m	100m	200m
101 original points	points	131	119	79	42
	rms %	2·1	2·3	2·5	4·1
25km^{-2} point density	peak %	-11	-12	-18	-24

Table a1.3 Results for the Reducing Area Experiment

interpolation factor = 3	measure	original	50m	100m	200m
selstr1	points	16	21	15	8
	rms %	12·1	12·4	12·5	10·5
	peak %	-40	-40	-40	-30

interpolation factor = 5	mesure	original	50m	100m	200m
selstr2	points	27	36	27	14
	rms %	9·1	9·9	9·4	9·2
	peak %	-35	-35	-30	-28

	measure	original	50m	100m	200m
selstr3	points	20	18	9	5
	rms %	18·4	18·5	15·1	13·9
	peak %	±40	±45	-40	±30

	measure	original	50m	100m	200m
selstr4	points	51	75	52	30
	rms %	4·8	5·4	4·8	6·1
	peak %	±20	25	±20	-25

	measure	original	50m	100m	200m
selstr5	points	37	56	36	23
	rms %	7·7	8·2	7·9	7·9
	peak %	-35	-40	-35	-35

interpolation factor = 3	measure	original	50m	100m	200m
selstr5	points	37	56	36	23
	rms %	9·2	9·5	9·3	9·1
	peak %	-40	-40	-40	-35

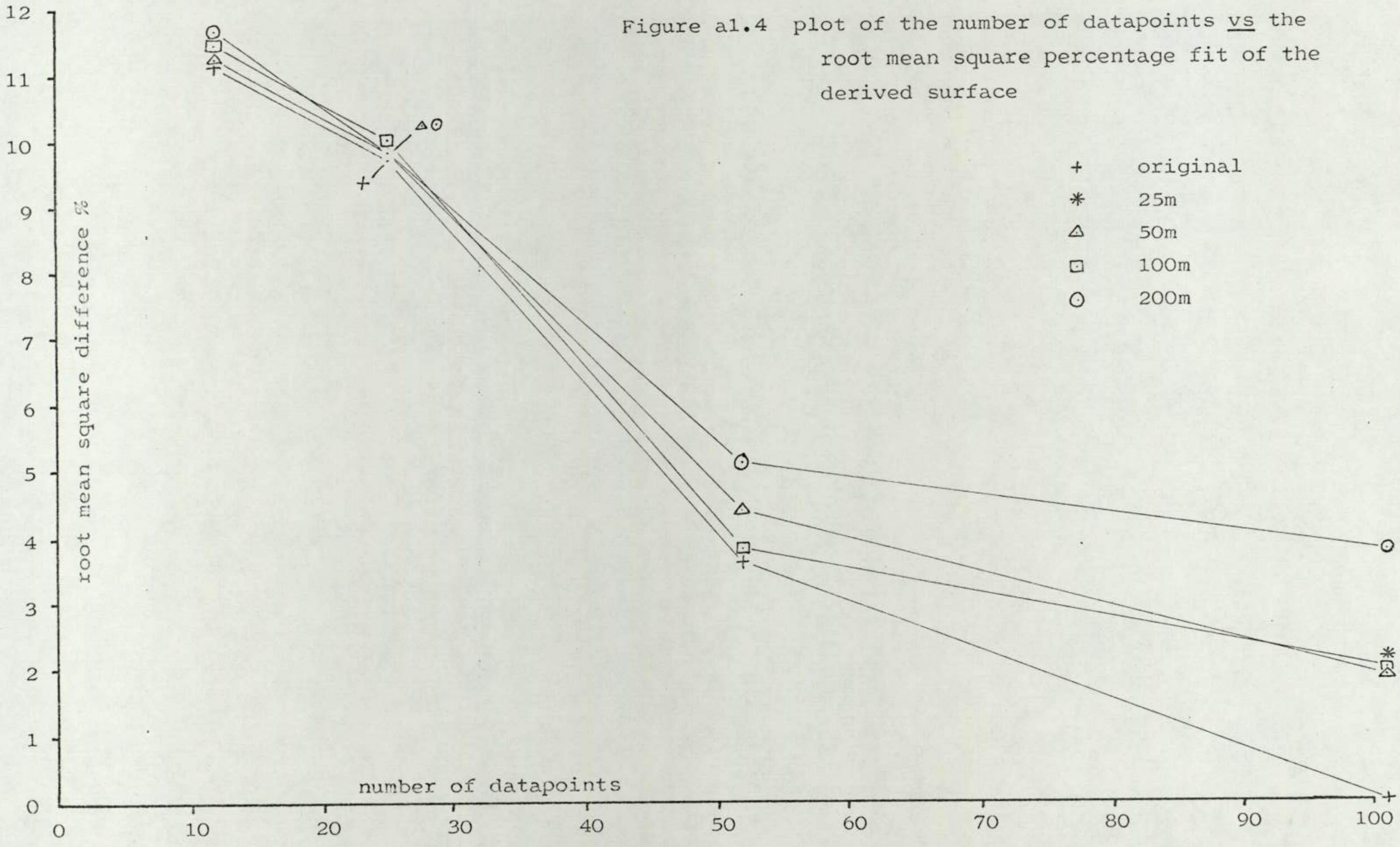


Figure a1.4 plot of the number of datapoints vs the root mean square percentage fit of the derived surface

However, tests in which the interpolation factor is reduced from five or six to three produce uniformly greater rms errors between the original and derived surfaces. This suggests that the pre-averaging should not be considered as part of the interpolation process and, implicitly, the distribution of the information below the threshold of the averaging grid is not only inaccessible but in fact largely irrelevant to later operations.

a1.1.5 General conclusions

It must be remembered that any of the interpolation processes used are being stretched beyond their original design capacity. They are intended for use where the mean separation of the datapoints is <u>less</u> than the pixel size and, in the case of the large mesh sizes in this experiment, are performing the additional function of filling in intervening points to allow comparison with other DTMs to be made on a similar basis.

From this study it is clear that there are a number of considerations when choosing a pixel size. We are faced with the odd association of detailed, dense, efficient site investigations, optimal single points from trial pits or scientific investigation boreholes and <u>no</u> intervening data. In effect there are two mapping scales on one map.

In this project the choice has been made to average the higher-density areas to provide a better basis for interpolation and a method of defining the many parameters for the whole map in a reasonable size of database. The specific requirements of the project are as follows:

Size of the Final Database. Obviously the smaller the better, concomitant with the resolution of the final map, bearing in mind the enormous increase in the storage requirements for each reduction in pixel size. e.g.:

pixel mesh size in metres	area in pixels	number of pixels in the test area
10	200X200	40,000
25	80X80	6,400
50	40X40	1,600
100	20X20	400
200	10X10	100

Input Data. These have been analysed above. Interpolation programs produce a peneplane effect in areas with no data. This is found to be more reliable and realistic for geological data than the drastic effects of the mathematical extrapolation programs which are commercially available. From examination of the plots from the reduced data area experiments it may be seen that an increased pixel size 'spreads' the data and appears to provide a longer 'reach' into the unspecified areas than smaller pixel sizes.

Amplitude and Wavelength of Errors. From the data reduction experiments we find that the rms goodness of fit is more affected by the large wavelength rounding errors than by sharp edge-effect peaks. This indicates that overall a smaller, circa 25m, pixel would give a better representation than the very large (200m) pixels, when the data distribution is dense and random.

The Nvyquist Wavelength of the Collecting Pattern. This is implied in the pixel mesh and leads to the following minimum resolvable features:

pixel size in metres	Nvyquist wavelength in metres
10	25
25	62·5
50	125

(cont.)

pixel size	Nvyquist
100	250
200	500

Reference should be made to Figure a1.3, the map of data distribution. Whilst a resolution of 250m may seem unacceptably large as far as the plotting of streams or gravel terraces is concerned it is little different in effect from the 125m of a 50m pixel, and significantly more useful than the statistically significant 500m of a 200m pixel. However, the empirical performance of the 100m pixel on real test data in the calculated comparison tests is acceptable. This is probably due to a fortuitous balance between the reduction of cluster effects, shared point and rounding effects. Interpoint filling of the mapped matrix at a ratio of two points for one during the calculated comparison experiments reveals that, if required, the surface may be 'drawn out' to a greater effective resolution than expected. It must be pointed out, however, that features below 250m in wavelength must be viewed with some caution, even though the experiments performed show only rounding errors.

Accuracy and Location of Retrieved Features. Shared point effects and translocations are only severe with small pixels in areas of high information density. It should be recorded that some shared point effects were noted with very low point densities at pixel sizes up to 100m.

Estimation of Errors. The number of points before and after averaging provides an empirical guide to the influence of the shared point effect. A simple rule is that, if there are the same number or more points after averaging than before, the quality of the surface will have significantly deteriorated from the original. During interpolation an estimate of the quality of the interpolation may be made from: the number of points, the interpolation factor and the mean separation of the datapoints.

a1.2 Determination Of The Optimum Interpolation Function

a1.2.1 Introduction

For a simple interpolation routine the influence of the various points included in the average is determined solely by a weighting equation, the INTERPOLATION FUNCTION. This function determines the realism of the interpolated surface. Shep~~p~~ard (1968) proposed the function used by the Ordnance Survey: $\frac{(1-D)^2}{D^2}$ where the radius of influence $D<1$. Davis (1973) suggests that a $\frac{1}{\text{separation}}$ weighting is suitable for most data distributions.

Two sets of experiments were carried out to ascertain the most appropriate interpolation function for the distribution provided by urban site investigation data. The choice was made on a subjective appraisal of sets of interpolated surfaces (the third dimension described as Z) produced in the two tests.

a1.2.2 Methods

A totally subjective test, employing the DPLOT program described in Chapter 5, was carried out first. For this test three sets of data were interpolated. These were: a peak in the centre of the map with the edges tied down to a lower value (9 points), an undulating edge to the map (8 points) and a uniform slope (9 points). The Ginosurf function RANGRD was employed as a reference function. RANGRD is Shepard's (1968) routine followed by a second order smoothing process. The functions tested were: $\frac{1}{d}$, $\frac{1}{d^2}$, $\frac{1}{d^3}$, $\frac{1}{d^4}$, $\frac{1}{e^d}$ where d is the separation.

The second test employed a variant of the program used in the mathematical comparison of the different pixel sizes. The presentation and calculations were the same as in the earlier program but the various pixel size surfaces were replaced by the functions: $\frac{1}{d}$, $\frac{1}{d^2}$, RANGRD and $\frac{(1-D)^2}{D^2}$. The reference surface is interpolated with a $\frac{1}{d^2}$ weighting because this is the most efficient to use and because it performed well in the first test. The test data were the same as for the mathematical comparison tests given in a1.1. These offer a 14% areal coverage with 50m pixels and 25% with 100m pixels. The results were analysed in the same manner as the data from the pixel comparison test.

a1.2.3 Results and conclusions

The first test proved a direct relationship between the power of the distance in the weighting function and the shape of the calculated surface. The greater the power of the factor the more the closest point is allowed to affect the calculated Z value. Thus the lower powers produced the most 'spikey' representations. The unity and square factors developed the closest approximations to the RANGRD surface. In fact, in the eight datapoint test the $\frac{1}{d^2}$ surface appeared more realistic than the Ginosurf plot.

The second test quickly became a study of the usability rather than the accuracy of the interpolation factors. RANGRD will not function on less than six data points and has a number of array and input requirements which make it difficult to employ in a flexible fashion. For the proposed application this route is therefore unusable and was dropped from consideration. However, as a reference RANGRD was retained in the test program output.

The Shepard function, $\frac{(1-D)^2}{D^2}$, takes its datapoints from a pre-specified radius of influence. For a good calculated surface, with the range of influence minimised, this radius should be as small as possible. The radius is automatically increased by the program if insufficient data points are available for the current point. A normalised average value for the radius achieved by the interpolation provides an indication of the suitability of the chosen initial input value for the data distribution provided. This value was named the move out correction of the function.

If the move out is not unity the function has not interpolated the surface uniformly. The resultant switching between two or more radii of influence, with concomitantly different sets of initial points, will cause an increase in the noise on the calculated surface. Tests showed that the data distribution of the site investigation results used in the tests (NB taken from the 'best' of Varey's data) was too sparse for this function to be employed in a satisfactory manner. In addition, the specification of a particular radius of influence to suit the dataset was considered to be too sophisticated for use in the final map system.

The Shepard function was, however, employed for the spreading of the information taken from the O.S. 1:10,000 map which was to act as a present day DTM for the database. In this application it functioned well, producing a very detailed surface. The initial data for this process offered a coverage greater than 50% of the mapped area; while the site investigation information for the final database map has only an 18% areal coverage.

Between the remaining $\frac{1}{d}$ and $\frac{1}{d}2$ functions all the experiments indicate that the $\frac{1}{d}2$ weighting offers a much more realistic approximation to a natural, continuous surface with the test data used. $\frac{1}{d}2$ has therefore been adopted for the real variable interpolation function in the map database system.

a2 CODES AND FORMATS FOR THE MAP DATABASE

a2.1 Numeric Codes used in the Datastore (with their respective frequencies of use)

a2.1.1 colour

code	column 1	column 2	column 3
1		pink-	red
2	light	red-	yellow
3	dark	yellow-	brown
4		brown-	olive
5		olive-	green
6		green-	blue
7		blue-	white
8		grey-	grey
9		orange	black

a2.1.2 qualifier for the second colour

code	parameter	original	imperial	metric
1	mottles of	30	13	6
2	laminations of	30	2	14
3	vertical striations of			
4	spotted with	11		
5	grading into	11	12	11
6	and	164	43	150
7	lenses of	2	8	10
8	traces of		8	52

a2.1.3 weathering grade

code	parameter	original	imperial	metric
1	fresh	41		
2	faintly weathered	43		
3	slightly weathered	77		
4	moderately weathered	65		
5	highly weathered	26		
6	completely weathered	9	1	
7	residual soil		38	32
8	'weathered'	1	77	174

a2.1.4 stratigraphy

code	parameter	original	imperial	metric
00	Recent			
1	fossil soil		7	
2	topsoil	48	112	63
3	alluvium	202	73	180
4	head			
5	brickearth			
6	river terrace	1	19	8
10	Glacial			
11	undifferentiated drift	697	385	778
12	Older drift		1	
13	Younger drift			
14	Hoxnian			
15	interglacial		3	
20	Made			
21	made ground	503	239	702
22	chemical	6		
23	organic			1
30	undefined		134	215
40	Triassic			
41	Arden Sandstone			
42	Keuper Marl	356	147	262
43	Lower Keuper Sandstone	385	36	56
44	Upper Mottled Sandstone	26	30	27
45	Bunter Pebble Beds			3
46	Hopwas Breccia			
50	Carboniferous			
51	Coal Measures			

a2.1.5 single number stratigraphies for stored maps

code	parameter
1	interglacial deposits
2	Keuper Marl
3	Lower Keuper Sandstone
4	Upper Mottled Sandstone
5	Bunter Pebble Beds
6	Boulder Clay and Undifferentiated drift
7	Sands and Gravels
8	alluvium
9	river terrace

a2.1.6 lithology

codes	parameter	original	imperial	metric	text
1	breccia				
2	conglomerate	52	1		0013
3	coarse grit			1	0013
4	medium grit	381	1		0013
5	fine grit				
6	coarse sandstone		10	4	0040
7	medium sandstone	225	11	29	0040
8	fine sandstone		5	7	0040
9	siltstone / shale	27	8	20	0400
10	mudstone / clay	245			4000
11	greywacke	52			1111
12	coarse limestone				
13	medium limestone				
14	fine limestone		1		
15	pebbly sandstone	70			0031
16	marl	260	70	34	1300
17	marl & siltstone bands		30	37	
18	sandstone & marl bands		4	8	
19	limestone-shale		2		
21	soil				
22	sedimentary	78			
27	rock	6			
31	very coarse gravel		1	5	0013
32	concrete gravel	7	39	60	0022
33	washable clayey gravel		10	26	0022
34	clayey gravel	84	15	9	1012
35	gravel-sand		106	271	0031
36	gravelly clay	108	20	29	1101
37	sand-clay-gravel		90	240	1121
38	sandy gravel			13	0013
39	pebbly sand	1	1		0013
40					
41	sand	41	105	129	0400
42	silty sand	88	35	51	0130
43	sandy silt	4	10	2	0220
44	silt		13	2	0400
45	slightly clayey sand		8	36	0130
46	clayey sand	42	20	66	1130
47	clayey sandy silt		9	4	1110
48	clayey silt		33	253	1210
49	sand-clay	2	3	2	1120
50	sand-silt-clay		52	29	1210
51	silt-clay	1	15	8	1300
52	sandy clay	76	34	9	1110
53	silty clay	6	45	185	1200
54	clay		57	13	4000
55	sand & clay bands		1		

The texture (text) is the proportion, to the nearest quarter, of clay, silt, sand, gravel in the lithology type. The equivalents for the made and mixed types are given in Plate 5.4 .

a2.1.6 cont.

code	parameter	original	imperial	metric	qual
58	sludge			1	19
61	rubble	7	7	91	3
62	fine fill		38	66	1
63	ash		14	33	13
64	sludges		1		19
65	timber & bricks			5	8
66	wood,metal,glass,ash etc.		1	7	9
67	steel & concrete		1	1	7
68	ash & coarse fill		37	164	11
69	building waste			5	6
70	soil		1	1	4
71	concrete		3	5	5
72	ash and fine fill		50	51	10
73	rubble & fine fill		33	198	2
74	'fill'	519	36	43	-1
75	cinders		3		15
76	old road		1	3	0
77	industrial waste			2	19
78	refuse		2	1	16
79	tar/flowing petroleum		1	5	17
80	slag			1	12
81	ash and slag		3	2	14
90	peat		7		
91	sandy peat				
92	sandy clayey peat		1	3	
93	clayey peat			1	
94	rotted leaves (pond)		1		

Qual is a score of the relative hazard associated with the particular type of made ground. It was used in the estimation of the 'worst made' for each segment.

a2.1.7 archive codes

code	parameter
1	1731 moats and pools
2	1825 moats and pools
3	first series O.S. moats and pools
4	first series O.S. tips and pits
5	recent refuse tips
6	recent brick and sand pits
7	toxic soils
8	private refuse / incinerators / dumps
9	industrial tips and transport sites

a2.1.8 qualitative strength

code	parameter	test results		orig	imp	met
	rock gradings	uniaxial mNm^{-2}	point kNm^{-2}			
1	extremely stg.	> 200	> 12000			1
2	very strong	100 - 200	6000 - 12000			15
3	strong	50 - 100	3000 - 6000			
4	mod. strong	12½ - 50	750 - 3000			
5	mod. weak	5 - 12½	300 - 750			
6	weak	1¼ - 5	75 - 300		6	2
7	very weak	< 1¼	< 75		2	5
8	indurated				33	16
9	strongly cemented					3
10	soft and hard bands				5	9
	soil gradings					
	coarse grained	S P T 30cm	50mm peg blows			
11	weakly cemented*				4	23
12	compact	> 10	< 100		15	39
13	loose	4 - 10	easy	340	55	165
14	very loose	< 4		20	10	25
	fine grained	compressive strength kNm^{-2}				
21	hard	> 288		176	44	48
22	stiff	144 - 288		555	50	33
23	firm	72 - 144		289	69	129
24	soft	36 - 72		290	53	45
25	very soft	< 36		5	5	5
26	friable				4	11
	organic layers					
31	compressed	firm				
32	compressable	spongy, open				
33	plastic	mouldable				
	* extra gradings	S P T				
35	very dense	> 50		303	69	131
36	dense	30 - 50		314	30	188
37	medium dense	10 - 30		272	55	591

a2.1.9 qualifier for the special feature or mineral

code	parameter	original	imperial	metric
1	rare (1%)	300	10	
2	scattered (10%)	238	1	42
3	common (30%)	232	24	17
4	abundant (50%)	628		
5	secondary			
6	matrix of		2	3
7	traces of		10	40
8	and		23	33

a2.1.10 mineral or special feature

code	parameter	original	imperial	metric
1	quartz	15		
2	mica	293	6	34
3	clay minerals	20		
4	calcite			
5	gypsum	25		
6	quartzite			
7	carbonates / calcareous		8	
8	post glacial organics	40	4	1
9	glacial organics	383	3	
10	recent organics	65	7	4
20	coal	401	2	4
21	cherts			
22	calcareous material	157		
27	sand and gravel		1	12
28	sand			8
29	gravel			20
30	bricks and timber			2
31	bricks			3
32	sand and clay			1
33	silt			3
34	broken shale			7
35	clay			12
36	cobbles		4	17
37	faulting		1	
38	limestone			

Figure a2.1
Coarse soil classification

after Varey (1977)

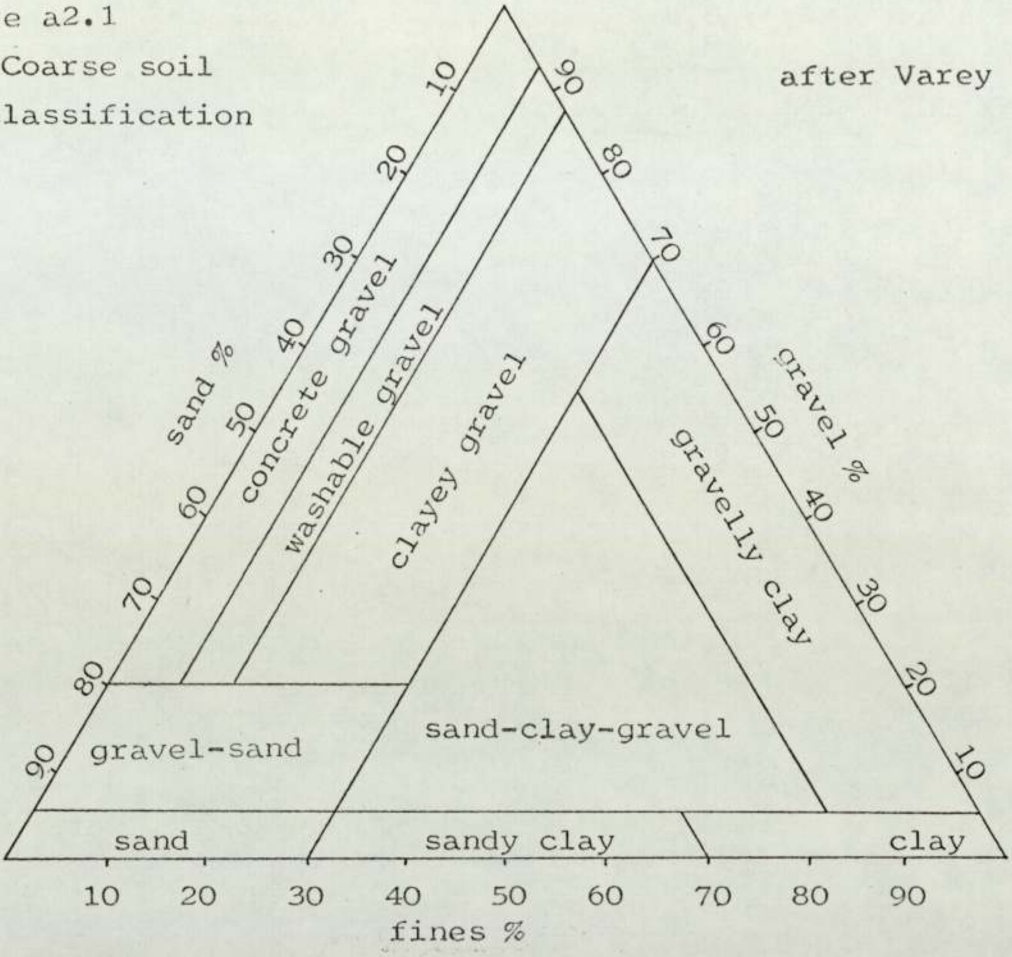
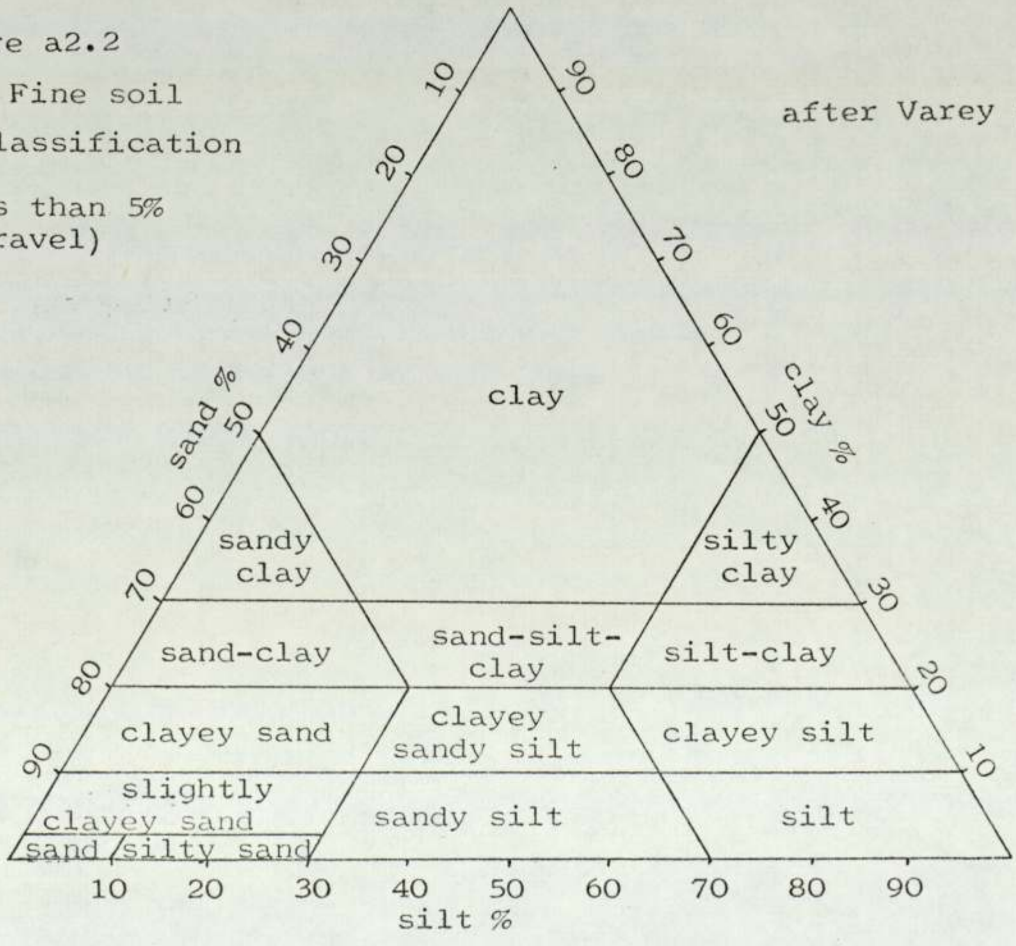


Figure a2.2
Fine soil classification
(less than 5% gravel)

after Varey (1977)



a2.2 Specimen Input Forms and the Special Codes used

If no surface level is given for the borehole the levels are entered as negative depths.

Header codes are altered in the following ways.

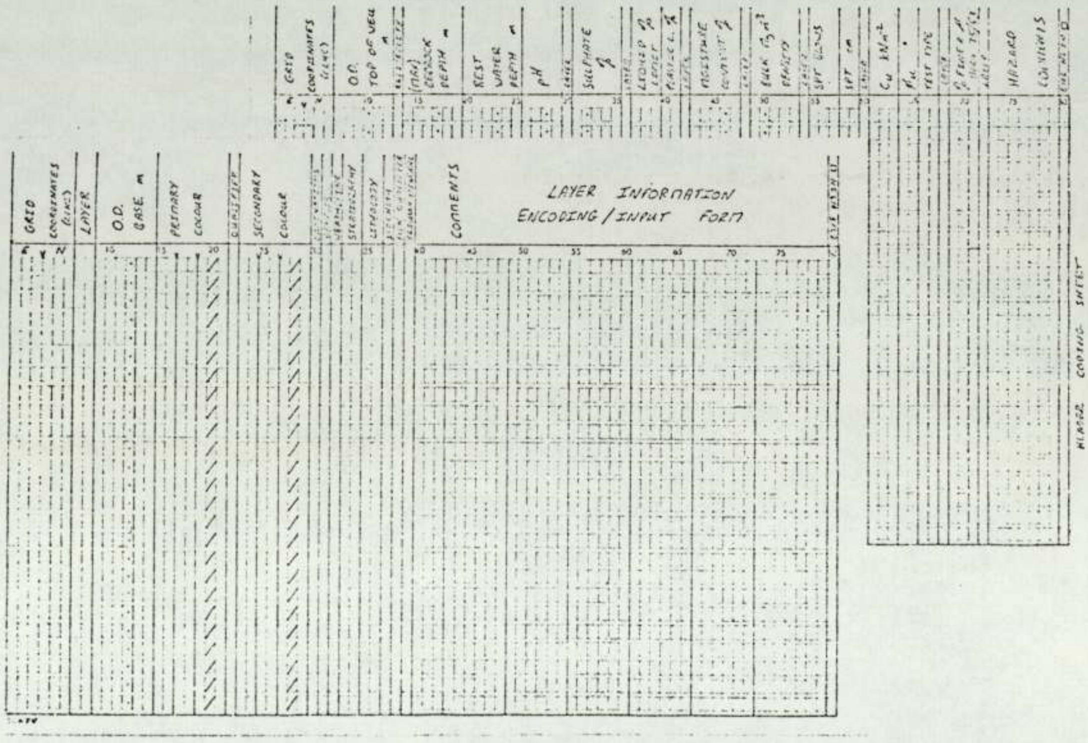
layers greater then ten are indicated by adding:

D/10, where D is the layer number, for the moisture content and shear strength

D/10+5 for the Atterberg limits, bulk density, pH and the sulphate concentration

if many values are averaged onto the form to save space the following format is used:

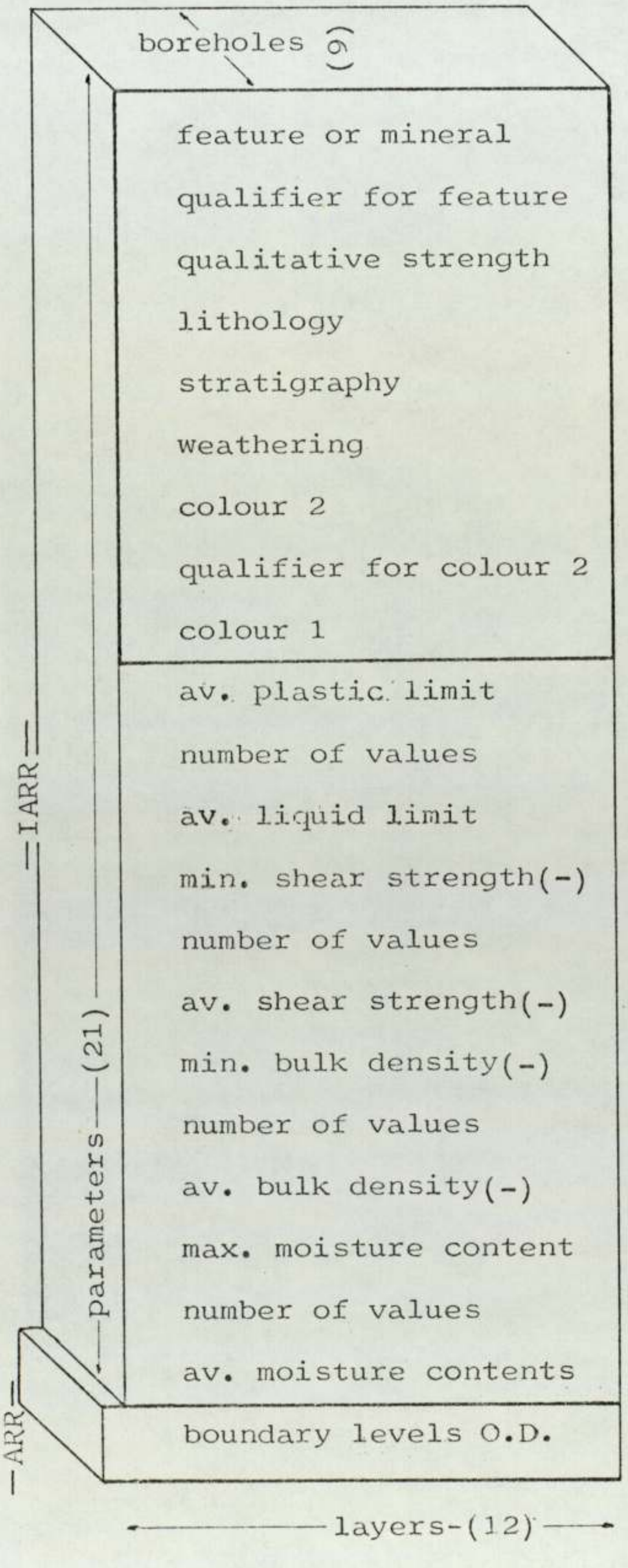
single	D/10	X	Y		Z	A	
value				•			source layer = C
average	A	X	Y		Z	B	= XY•Z
extreme	D/10	x	y		z	C	= xy•z
							number of values =(A-1)*10+B

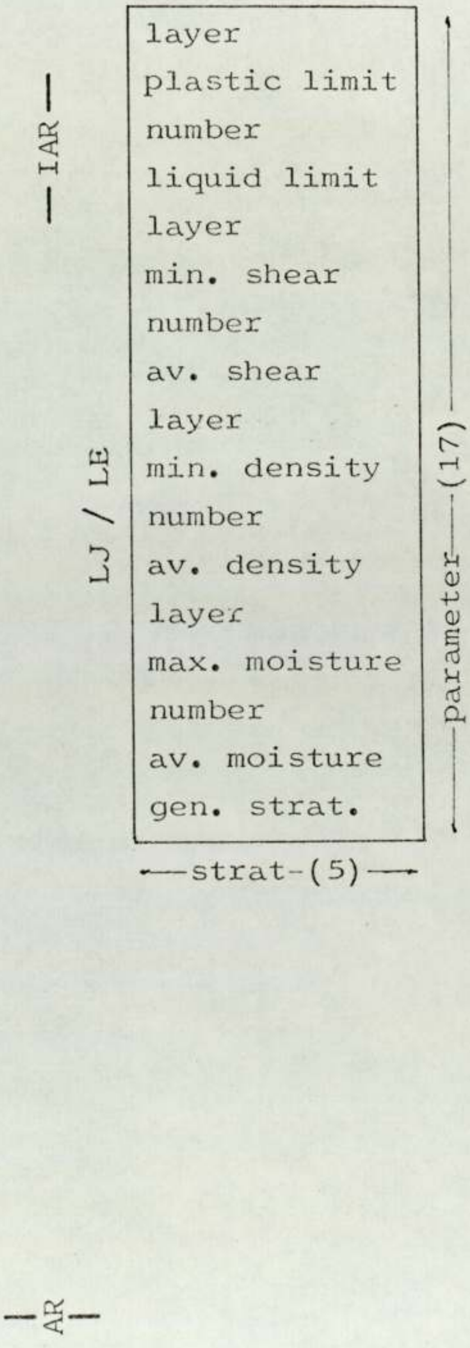


a2.3 Format of the pre-Processing files

```
line
  1   Xcoord Ycoord  surface OD  reliability  estimated deepest bedrock  rest water
         3      3       3·2          1                  3·1                  3·1

  2   number of pH readings   n( pH / layer )     (if  no test data n = 88)
                *                 *      *

  3   number of sulphate readings  n( sulphate / layer )    (if  reading is from water lays=0)
                 *                       *          *

  4   number of Atterberg tests  n( liquid / plastic / layer )  (if no engineering data  n=88)
                 *                    3         2        *

  5   number of moisture contents  n( moisture / layer )
                  *                      *          *

+ 6   number of averaged moistures  n( average / number / extreme / layer )
                  *                      *         *         *        *

  7   number of bulk densities  n( density / layer )
               *                      *        *

+ 8   number of averaged bulk densities  n( average / number / extreme / layer )
                     *                        *         *         *        *

  9   number of  standard penetration tests  n( SPT / layer )    (blows per 30cm is assumed)
                        *                        *      *

 10   number of shear tests  n( Cu / ϕ / layer )
               *                 *   *     *

+11   number of averaged shear tests  n( average Cu / number / extreme / layer )  (only if Øu<5°)
                  *                          *            *         *        *

 12-n layer number   base OD  colour1 q  col2  weath.  strat.  lithology  stren.  q  feature
           2           3·1       3    1    3     1       2         2        2     1     2

field widths indicated by subord. numerals  *indicates free format   + not in original data file
```

a2.4 Format of the Internal Data Buffers





a2.5 Format of the Final Direct Access Files

a2.5.1 Master or Header File

contents		field width	word
National Grid Reference	X	3	1
	Y	3	
number of layers		2	2
record number of first layer		5	
number of engineering records		2	3
record number of first set		5	
presence of a fault		1	4
archive hazard		1	
Geological Survey	Drift	1	
	Bedrock	1	
surface Ordnance Datum		4	5
proximity or number of boreholes		2	6
top of 'well' O.D.		4	
prox. or number reaching bedrock		2	7
original lowest bedrock O.D.		4	
prox. or number of water readings		2	8
average rest water level O.D.		4	
prox. ± and worst made type		2	9
original lowest made level O.D.		4	
worst pH value		3	10
prox. or layer number (0 if water)		2	
worst sulphate concentration		4	11
prox. or layer number (0 if water)		2	
worst Standard Penetration result		3	12
prox. or layer		2	
worst shear angle		2	13
layer		2	
existance map for the	Made	1	14
	Recent	1	
	Glacial	1	
interpolated	Drift map	1	
	Bedrock map	1	
prox (+6) map for the hazard map		1	
Hazard map		1	

a2.5.2 Engineering Data File

contents	field width	word
general stratigraphy	1	1
check coordinate	6	
prox. or number of values	2	2
worst average moisture content	4	
prox. or number of values	2	3
worst average bulk density	4	
prox. or number of values	2	4
worst average shear strength	4	
prox. or number of values	2	5
worst average liquid limit	4	
source layer for values	2	6
highest moisture content	4	
source layer for values	2	7
lowest bulk density	4	
source layer for values	2	8
lowest shear strength	4	
source layer for values	2	9
plastic limit	4	

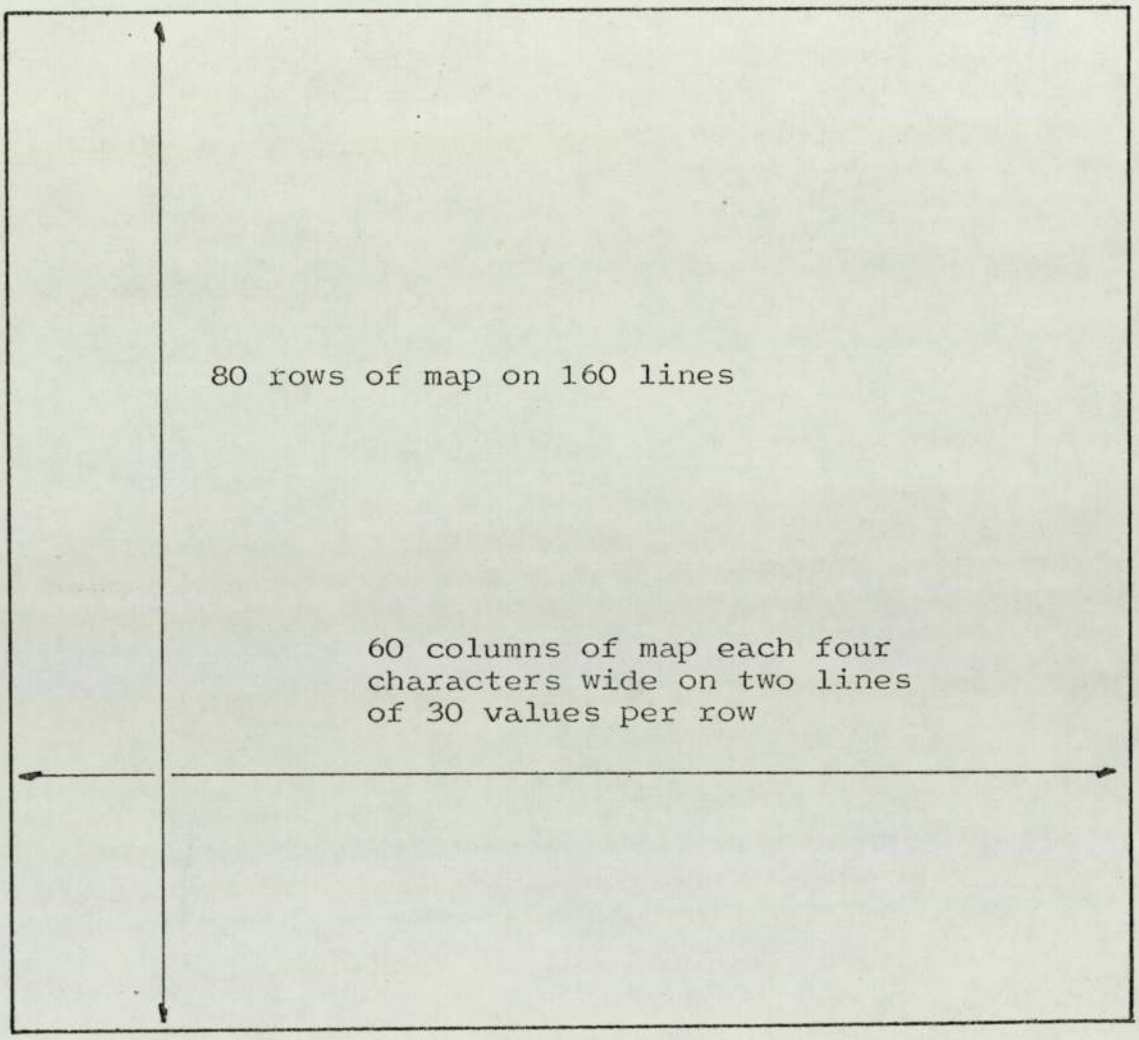
a2.5.3 Geological Data File

contents	field width	word
check coordinate	6	1
layer base level O.D.	4	2
weathering state	1	
qualitative strength	2	
primary colour	3	3
qualifier for secondary colour	1	
secondary colour	3	
specific stratigraphy	2	4
lithology	2	
qualifier for the feature	1	
special feature or mineral	2	

All real numbers are integerised to save space.
All codes are right-justified in their particular word.

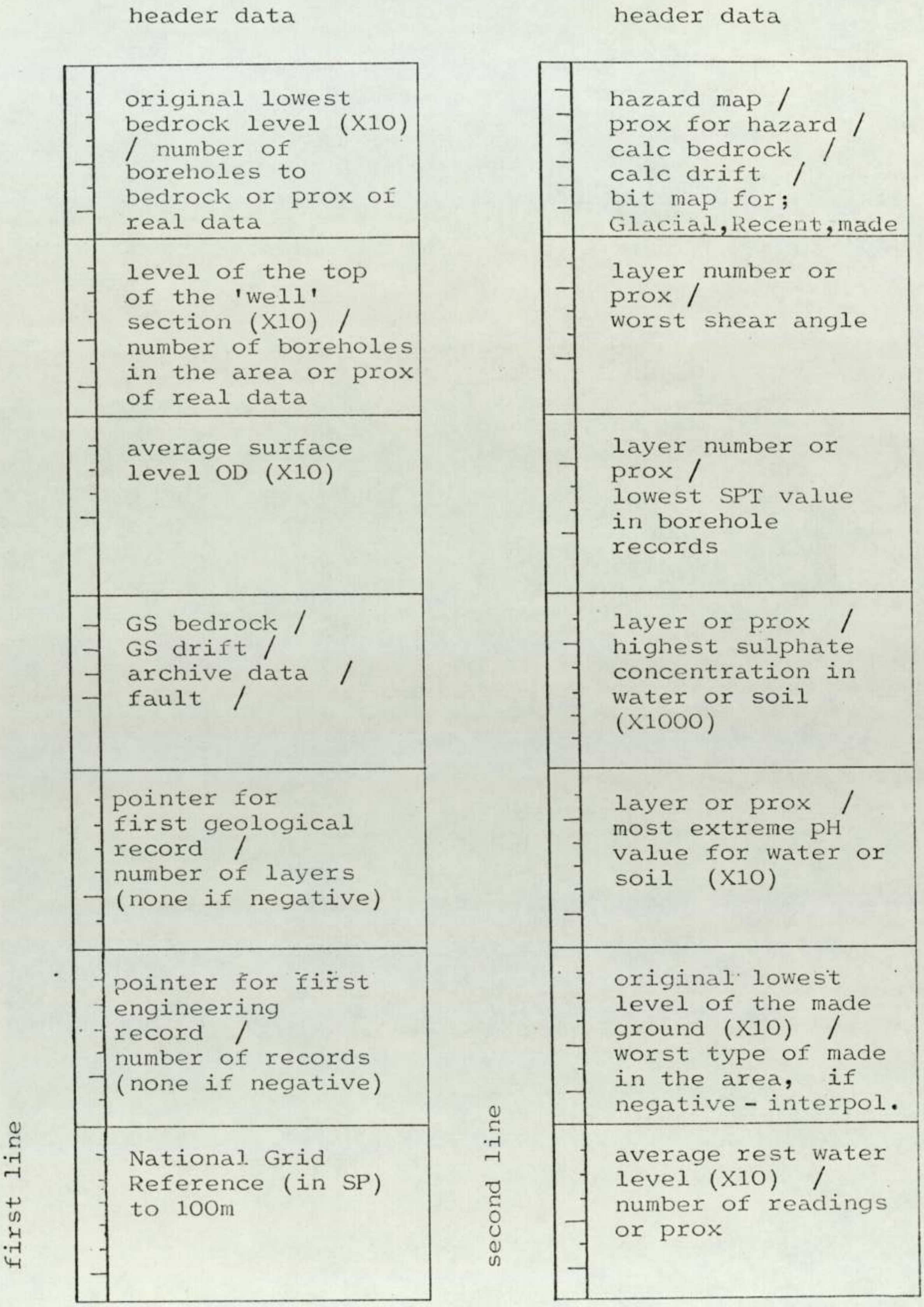
a2.6 Format of the Map Files

Scale Factor	File Type
real, free format	integer, free format
multiplication factor for the array values	0 - integer values 1 - classified values 2 - real values



Scale Factor for Classes	Number of Classes	n(integerised Class Boundaries)
real, 8·3	integer, 3	integer, 16 of 6 each

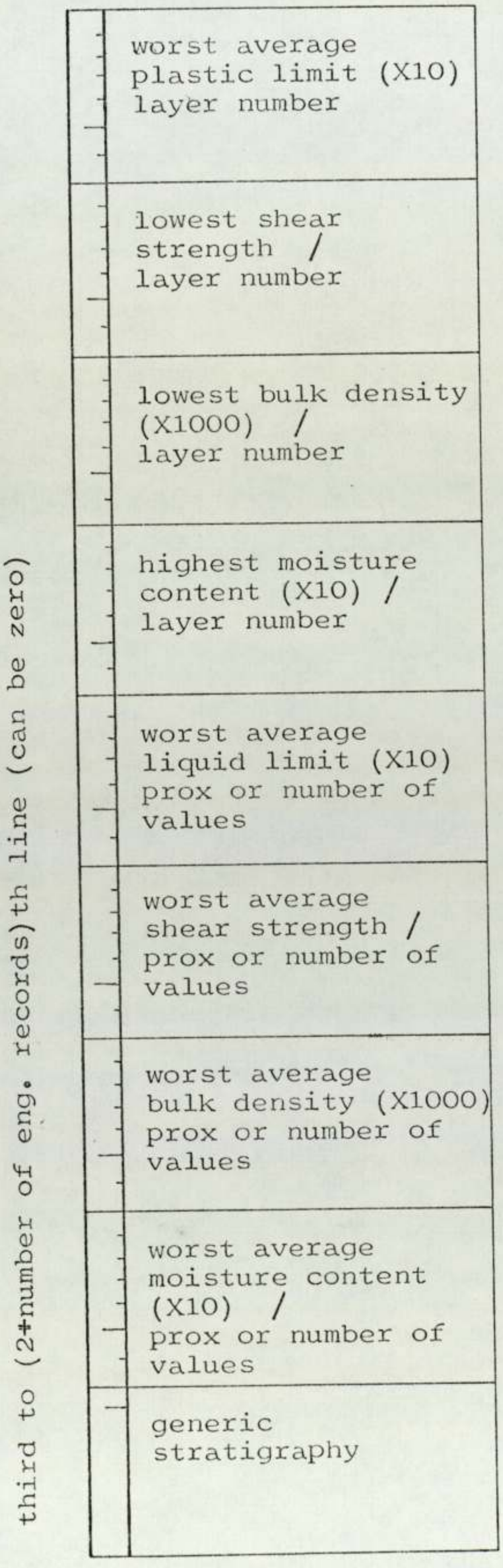
a2.7 Format of the Serial Listings from Extract



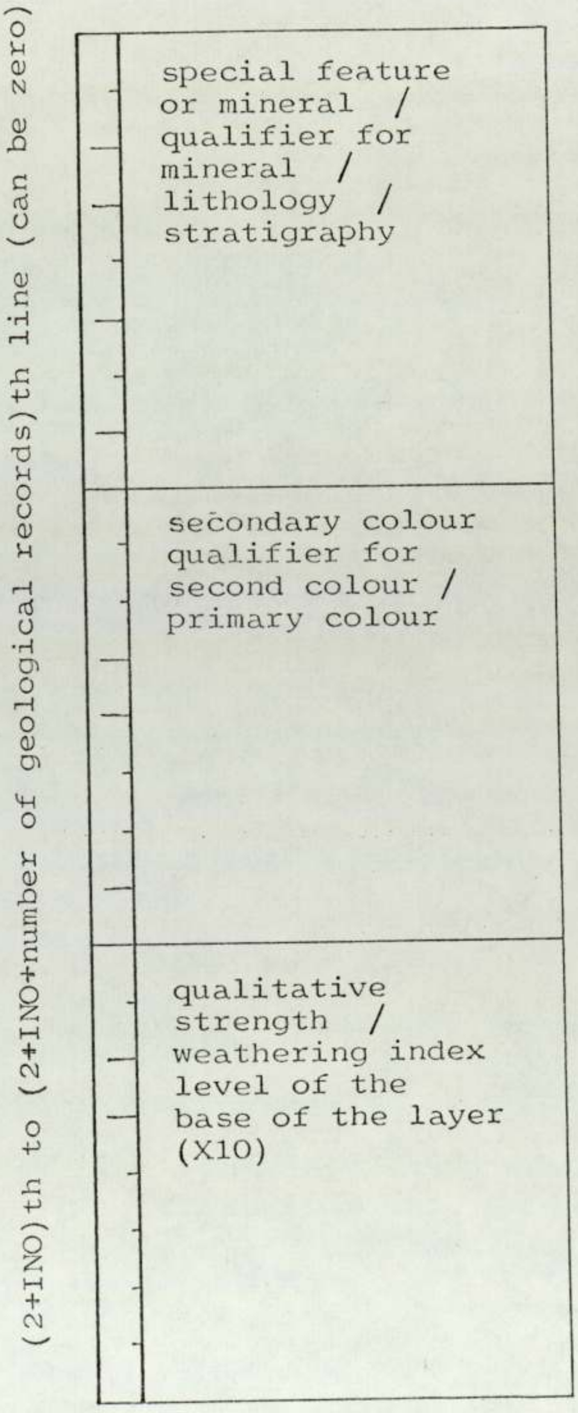
header data	header data
original lowest bedrock level (X10) / number of boreholes to bedrock or prox of real data	hazard map / prox for hazard / calc bedrock / calc drift / bit map for; Glacial,Recent,made
level of the top of the 'well' section (X10) / number of boreholes in the area or prox of real data	layer number or prox / worst shear angle
average surface level OD (X10)	layer number or prox / lowest SPT value in borehole records
GS bedrock / GS drift / archive data / fault /	layer or prox / highest sulphate concentration in water or soil (X1000)
pointer for first geological record / number of layers (none if negative)	layer or prox / most extreme pH value for water or soil (X10)
pointer for first engineering record / number of records (none if negative)	original lowest level of the made ground (X10) / worst type of made in the area, if negative - interpol.
National Grid Reference (in SP) to 100m	average rest water level (X10) / number of readings or prox

a2.7 continued

engineering data



geological data



a3 PROGRAM ALGORITHMS

a3.1 Introduction And Standards

The major algorithms developed for the database are listed in this section. A synthetic language is employed for this logic-table listing. It is hoped that the language is freely convertable into any logical programming language. The functions used are as follows:

ARRAY	description of an array. Specific forms used are: BUFFER for use in reading in and out of a routine RECORD a direct access read/write buffer REGISTER a look-up array loaded at runtime STORE a register which is permanently defined
CALC	any set of self-evident mathematical operations
CALL	JUMP to a subroutine or procedure
CH	channel number of a file or peripheral
DA	a direct access file
DECODE	a mathematical operation to split up packed or coded data also CODE
DISPLAY	write to a terminal
ELECTIVE	REQUESTed value used to branch a program by JUMPs or CALLs
ELEMENT	the current address of an ARRAY
IF	logical or mathematical comparison followed by a list of possible actions
INT	an integer number
JUMP	move to another part of the routine depending on some given condition
LIST	LOOP WRITE out a BUFFER
LOOP	continue to perform a set of instructions until some condition is satisfied
NEXT	pass to the next value in a loop
POINTER	an associated variable for a file or ARRAY
READ	the insertion of data from a source other than the terminal, usually includes a channel number
REAL	a floating point number
REQUEST	an interactive prompt for and READ in of data
RETURN	leave the current subroutine for the CALLing routine

SET	give a value to a variable, specific forms are: ASSIGN SET an array element LOOKUP value obtained from a register or store SAVE place value in a register array or variable PUT transfer a value to a new variable or array & PLACE INCREMENT used with variables in looped operations
WRITE	transmission of data to a device other than the terminal, usually includes a channel number
'N'	a generic counter integer
'X' & 'Y'	map or screen sense reals or integers for scanning or location

Special Graphics Commands:

CURSOR	interrogate the graphics cursor for its position
DISPLAY	output some graphical data e.g. text, figures, lines, a cursor etc.
DRAW	make a line from the previous to the new PEN position
MOVE	command to change the position of the PEN
PEN	current point of interest to the computer
SET	change some state in the graphics output device connected to the program

Most of these functions should be self-evident in their use. Reference should be made to the programs fiche in Appendix 4 to see the FORTRAN versions if more detail is required than the broad representations of the logic of these routines.

a3.2 Combine

Scans through up to four map arrays and their associated proximity matrices. Processes the values for each individual element by one of the built-in functions or a custom function edited in before compilation. Outputs one or two maps and a worst-case proximity map.

arrays

IAR & JAR	read & write buffers for maps, integer 60 wide
LAR & AR	arrays for maps and prox. rows, int, real, 60,6
PTS	register of scores for the difference option
FACT	register of scale factors for maps

logic

```
rewind all attached files
write proximity file specification to ch 12
display list of electives
request number of input maps, scale factor of output map,
     process elective and calculation constants (c1,c2)
```

```
if custom elective selected request if an extra map output is
     needed
if yes request specification of file
if second file needed write specification to ch 11
write specification of primary output file to ch 10 (if
     INVERT or CLASS options are specified mark as classified)
loop through input files and prox. maps
rewind and read specifications to FACT
end loop
loop 'y'
loop through data maps and prox. maps
loop read in rows of data to IAR & prox. values to LAR
if PERC option set sum score of points for each map to PTS
loop decode map integers to reals in AR
end loop
loop 'x'
loop select lowest or zero prox value for current element
     and set LAR(n,6) to value
elective call type of option: results assigned to AR(n,5&6)

DIFF   a*c1 + b*c2
PROD   a*c1 * b*c2
PERC   (a*c1)/b %  and score in c2
CINDEX (a-b)/(c-b)
CHOICE if a>b  opt1 = b & opt2 = 0 else opt1 = a & opt2 = 1
SELECT if a=c1 opt = b
GTEQ   if a≥c1 opt = c2
LTEQ   if a≤c1 opt = c2
INVERT if a=0 opt = c1-a+1  (reverses scores in class maps)
CLASS  if a=0 opt = a+5 and if opt>6 opt=6 (prox. to plotter)
CUSTOM user specified

calc integer value for output buffers
end loop
loop write rows of maps to ch 10&11(if req.) & prox. to ch 12
end loop
if PERC option calc mean of c2 from average of PTS scores
     display rms value
if inversion option write classes to map in reverse order
if CLASS write original values to map
end program
```

enhancements

Increase the efficiency, add more sophisticated proximity options, improve interaction - possibly via a limited language processor.

The proximity selection was altered in the following way for the hazard map:

Each map after the first is assigned a prox. factor. This factor is the minimum prox. integer this particular map's reliability is allowed to drop the overall reliability to. This provides a structure or weighting for the various maps when they are combined. The actual meaning has to be worked out at the time. For instance, if two maps of soil strength were being merged, one a widely spread qualitative index and the other of experimental shear test results, the tests would have a much poorer areal coverage. To avoid losing any information the test results should be given a -4 or -3 factor, while the qualitative information (possibly derived from the shear results, amongst other sources, in the first place) with an areal coverage almost certainly including that of the shear test results will be the master map with a factor of 0. Thus, the shear strength is used to enhance the qualitative strength values but is not allowed to affect the data distribution.

a3.3 Scatter

Divides the range of values in a map into 80 equal classes. Scores each class and displays the result to VDU or line printer (LP). Allows the division of the map values into up to 16 classes to give a meaningful or realistic classification via an interactive boundary selection system.

arrays

AR	map array matrix, real 60,80
CLASS	register of selected classes, real 16
IAR	input / output (I/O) buffer, int 60
JAR	screen buffer, score & class selection register, int 80,21

logic

```
rewind I/O chs 30,10
read & write map descriptions to chs 30,10
loop set JAR(n,1-20) to '.' characters
request & set 'null' switch
loop 'y'
loop read row of map to IAR from ch 30
loop through row, search for max. & min. values (if 'null' is
     set ignore zero values), decode elements into AR(x,y).
     count number of values
end loop
calc range of map values
loop 'y'
loop 'x'
calc scores for each of 80 ranges of values, store in
                                                  JAR(n,21)
end loops
INTERVAL = range/80
display max., min., scale factor & INTERVAL & write to LP
     file on ch 11
call SCALE (q.v.)
loop find max., min., scores & interval in JAR(n,21)
loop across JAR(1-80,n)
loop up JAR(1-JAR(n,21),1-20) setting elements to '*' charac-
                                                          ters
end loops
loop display JAR(1-80,1-20) to either VDU or LP file
request an 80 element row of ' ' or '1' characters for class
     intervals
loop read row into JAR(n,21)
loop across JAR(n,21)
if element = 1 calc value from INTERVAL & set next element of
     CLASS
end loop
request elective to integerise, repeat class selection,
     abandon or list to LP file or to continue, jump to
     appropriate statement
loop 'x'
loop 'y'
loop through CLASS to the range of AR(x,y)
```

```
assign 'n' to IAR(x)
end loop
loop out row to ch 10
end loop
write number of classes & loop scaled & integerised CLASS to
     the end of the map file
calc & write histogram vertical scale & number of values to
     LP file on ch 11
end program
request listing of LP file
```

enhancements

Automatic class selections for 16,10,6,4 ranges of values. Magnification of sections of the distribution. More sophisticated statistics. Transfer to graphics VDU with more display ranges and simultaneous display of the map locations during the build-up of the histogram. Cursor driven classification.

a3.4 Choro

Plots an isopleth matrix representation of a map file. Up to sixteen colours may be used, specified by a VLT file. The display may be 'sliced' for any of the ranges, with an indication of the values above & below these values. A line printer copy of the map can be plotted, in up to 16 default numeric or user-specified symbols.

arrays

IAR	read & write buffer, int 60
JAR	map matrix, int 60,80
ICLASS	data ranges register, int 16
CLASS	data ranges register, real 16
IVLT	VLT register, int 3,16
ITX	line printer character store, alphanumeric 16
IT	title buffer, alphanumeric 12

logic

```
rewind all files
read map specification
loop read VLT settings from ch 13 into IVLT & set graphics
                                                        VDU
call BASE (displays horizontal frame on VDU)
call SIDE (displays vertical frame on VDU)
loop 'y'
loop read row of map from ch 30 to IAR
loop set JAR(x,y) to IAR(x)
move pen to start a new row
display IAR on the VDU as a row of boxes coloured according
     to the value of the element and VLT
end loop
if file type is not 'classified' jump to end, print, slice
     options
move pen 'x'
read number of classes & loop read to ICLASS from ch 30
loop 'n'
calc real value & set CLASS(n)
```

```
display n & class value in current VLT colour 'n'
display default symbol for LP map from ITX(n)
increment 'y'
end loop
label class list & symbols
request elective to end run, plot map on LP, bit slice map
end program if selected
request title, loop read & display on VDU
if slice requested call SLICE
end if return elective from SLICE is 'end program'
request elective for specified or default symbols
if specified selection initiate cursor and loop position
     against values to be replaced, requesting replacements,
     display entered symbols in place of defaults on VDU &
     in ITX
loop write title to LP file ch 10
loop 'y'
if classified map loop write ITX(JAR(x,y)) to LP file ch 10
if unclassified map loop write JAR(x,y) to LP file ch 10
in both cases if 'y' array address' last digit is zero add
     'y' map coordinates to left & right of the displayed
     map, else spaces
end loop
write 'x' map coordinates to LP file ch 10
write labelled number of classes
loop write CLASS with a label
loop write ITX with a label
end program
```

subroutines

```
BASE
set pen 'x' & 'y'
loop botton & top of screen
loop 'x'
move pen 'x' & display eastings label
end loops
return

SIDE
as BASE transpose x & y

SLICE
request class intervals to be highlighted if value = -1 jump
     to reload request
loop set VLT addresses below value to a pale green
set selected value to yellow
loop set addresses above selected to pale red
jump to class request
request elective to return or reload original VLT
if no reload jump to end request
loop set VLT to values in IVLT
request end of program or print map elective
return
```

enhancements

Allow textures as well as colours on the graphics terminal. Permit vector overlays. More interaction in the special character selection.

a3.5 Section

Reads out a three parameter map west-east or south-north through the database map. The section is textured to the lithology, coloured and has an indication of the stratigraphy for each lithological layer. The section contains continuous linear representations of the level data held, plus one optional extra map file. A copy of the screen image may be made to the line printer, with a note of the qualitative strength in place of the colour of the soils. The presence of real records is indicated.

arrays

IMAP	map index array, int 60,80
ISTP	buffer for graphics information, alpha 45
JAR	extra map file register, int 60,80

logic

```
rewind all files
request elective for connected map file
if set loop load JAR and find max. & min. values
calc scale factors of map data for screen & LP displays (q.v.)
loop read map index into IMAP and read DA file parameters
switch on graphics
loop read symbol definitions and screen background from pre
     created file to set up the graphics terminal via ISTP
set background colour
define DA files
request elective for cursor driven input
calc elective to:
     show a new section, jump to request for a coordinate
     rescale the current section, jump to call ERASE
     draw the current section on the LP, jump to LP params.
     end the program
     call BACKGR
request the National Grid coordinate & electives for a new
     section orientation, number of elements
set scales, calc & label the current start coordinate
calc south-north or west-east start & finish coordinates on
     the map index array
check if the section will run off the edge of the map &
     truncate if necessary
call ERASE, call XSECT, call NSWE & if the rescale is in
     operation call ROWS
jump to cursor elective
request parameters for the LP listing width, spacing &
     archive record
set screen background to black
call START, call XSECT, call START, call STOP
set background to original
end program
request printout of sections in LP file
```

subroutines

XSECT

arrays

IA	register of archive data, int 4
J	buffer for DA records, int 11

AR	register for the levels O.D. in the well log, real 13
IROW	buffer of continuous level data, integer 80,5

logic

```
initialise variables
set up increment widths for the labels & the number of chara-
    cters for the section elements, limiting width at 12
    characters
loop set IROW to zero
calc scale factor for the screen or the LP for the section
    levels
loop across the section
increment the drawing position in 'x'
lookup current header address in the map index
lookup the current map file value (if selected)
read DA record
write the number of the section element in black or white
    depending on the presence of real data - the sign of J(3)
decode the continuous level data into IROW
decode the archive data into IA
select the max. & min. levels over the whole section for the
    continuous and well levels
if no geological data record, jump to level data output
if LP output elective in operation call PREBL
loop
read DA geological data record to J
decode colour, lithology and stratigraphy
calc top of layer & thickness on the screen
if LP output call LLAY & jump to level data else call DLAY
end loop
check AR for lowest level
if not LP output next else call LEVEL
if no data and 'proportional' elective is set write one
    segment of spaces to the LP file ch 10
else call LOUT
end loop
redetermine screen scale and end corrections from the current
    overall max. & min. levels O.D.
flush graphics buffer
return
```

ROWS

logic

```
loop across parameters in IROW
if  map file is selected set loop to 5 times
calc map scale factors
reset pen 'x' to the left hand side
set current pen colour
loop across section in 'x' and in IROW
calc absolute real value from IROW
if value is zero next
calc pen 'y' from absolute value
if last value was zero move to new 'x,y' else draw to that
    location
increment 'x'
end loops
flush graphics buffer
return
```

LEVEL

arrays

IDTM	LP buffer, alphanumeric 130
ITX	store of LP symbols, alpha 5

logic

```
call PREBL
loop through IROW 1-4
if value is out of range next
calc address equivalent to current O.D. level & set element
     in IDTM to ITX(n)
end loop
if a map is selected calc address, if out of range jump to
     loop out
set buffer element to map symbol IDTM('y') = ITX(5)
loop write IDTM to ch 10 with the selected archive code at
     the end of the string
return
```

PREBL

```
loop across and down the LP buffer, set all elements to ' '
character, return
```

NSWE

```
display a SCALE (q.v.) of current height range on the l.h.s.
     of the screen
display labels of W - E or S - N on the r.h.s & l.h.s. of
     the screen
electively display a SCALE of map file range on the r.h.s.
     of the screen
return
```

START

arrays

IND	store of labels, alpha 2,3
IBL	store of alpha numerals, alpha 10
ITX	LP buffer, alpha 7,130

logic

```
write a header of N.G. coordinate & the orientation, from
     IND, to ch 10
write a SCALE of the height range to the buffer
loop write the buffer to ch 10
return
```

STOP

```
as START for the scaling of the height but use the map file
write a table of  explanations & symbols for the section on
     paper to ch 10
return
```

LOUT

array

IBIT	buffer for label, integer 7

logic

```
set the current width extremities of IBIT to the number of the segment in the section and the number of original boreholes in the segment
loop out the buffer ITX to ch 10 with an element of IBIT added at the right hand edge of each row
return
```

LLAY

logic

makes up a buffer full of the current segment's geological information in LP characters. The operation is identical to DLAY (q.v.) except that the strength of each layer is substituted for its colour.

enhancements

Switchable parameters in place of the lithological texture and geological colourings in use at present. This would enable the engineering data to be meaningfully presented in the display. Additional texture & colour routines would be required for this simpler representation.

Pseudo-three dimensional display of sections behind and in front of the current section, to create greater interpretative realism.

a3.6 Segment

Program producing a translated verbal listing of a single segment of the database. The program will also give coded listings of the same data in the same format. A simple facility to insert new whole records for segments is provided, allowing comparison with the original data held and a checklist before electively replacing the old record. The replaced datas' pointer values are output to a file and the old record remains in the datastore until the next restore, or REPLACE, takes place.

arrays

IMAP	map index register, int 60,80
AR	register of levels O.D. in a segment's lithology, real 13
IAR	a register of the geological data codes for the current segment, int 12,9
LJ	a register of the engineering results for the current segment, int 18,5

logic

```
rewind all files
loop read map index to IMAP
read DA file parameters
set DA file parameters plus an increment to allow for inserted records
define DA files
call READTX
request elective for the operational mode:
    no translation
    translated records
```

```
     insertion of new data, no translation set by default
request the N.G. coordinates of the segment to be retrieved
calc map index element
reset program switches
if element is outside the map display a note & jump to coor-
     dinate request
call READ
if data insertion is selected request number of layers & eng-
     ineering records
display general header data
if inserting request replacement row & set relevant variables
if no further data is available for segment (switch IST from
     READ) jump to new coordinate request
display header data for known segments
if inserting request replacement row of data
display label for the eng. data
loop
display eng. records & stratigraphic horizon, if translating
     divide values by conversion factors & display as reals
if inserting request replacement row
end loop
display geological label & number of boreholes, level of top
     of well AR(1)
if inserting request replacement row (in integer from)
loop
if translating call TLAY
else display AR(n+1) and loop out IAR(n,1-9)
if inserting request & set replacements in coded form
end loop
if not inserting jump to request for listing
request elective to replace records or abandon the insertion
if abandon jump to request for listing
reposition pointers at end of DA files
call WRITE
update max DA pointer values & address for current map
     element
request elective for listing to an LP file, if yes repeat
     all display operations to ch 10 plus added line feeds
     at the end of the listing
request an elective to:
     continue in the same mode, jump to coordinate request
     change mode, jump to mode request
     end run
at end of run, if any insertion has been performed, loop out
     new map index & new DA parameters
end program
request listing of the LP file
```

subroutines

READTX

arrays

ITXT	store of text variables, alpha 1300 (3 char per word)
IRG	register of pointer & string lengths, int 2,3000
ITP	start pointers for different parameters in ITXT, int 11

logic

```
loop reads text types from a datafile with character lengths and equivalents for the numeric codes
put in an array ITXT, simultaneously making up a list of start (ITXT(n)) addresses and string lengths (n1-n2) in register IRG
whilst storing the (IRG(1,n)) addresses for the commencement of each new set of parameters in ITP
```

TLAY

arrays

IC	register for colour code elements, int 3
IEQ	store of parameter codes, int 9
JTX	output buffer for completed text descriptions, alpha 72,3 (A1)
IOUT	text string buffer for each parameter, alpha 36 (A1)

logic

```
loop set output array to ' ' character
loop through IAR
if parameter not specified next
lookup check for a colour (IEQ=1) if yes decode elements & loop the translation three times instead of one pass
call WTEXT which:
    lookup start address of current parameter in ITP
    lookup length & start address of text in IRG
    loop extract text from ITXT into IOUT (A3 to A1 trans.)
    return
call REPORT which:
    check if space is available on the current print line for the new string
    if not switch to new string & reposition the buffer pointer
    return
display the O.D. from AR and loop out the test buffer to the screen or write to ch 10
return
```

enhancements

Provide more translation and better format on the line printer to make use of the greater paper area. Softer data insertion with a 'batch' mode for the insertion of several segments. Automatic printout of the eight adjacent segments to that selected for convenience in site investigations. Automatic indication of the coordinates of the nearest hard data if segment is interpolated.

a3.7 DRAW

Uses a high level graphics library of routines (GINO-F) in the custom, interactive production of graph plotter hard copy. Plot types are Isometric, Isoline & Isopleth. The isopleth is not a package but is made up of lower level routines. It allows coloured, shaded maps with several types of shading & symbols to be plotted at the same scale as the isoline maps. The program enables graphical output to any device which has a GINO backend. This includes an intermediate code store for overlays or production runs.

arrays

LC	store of pen colours for the devices used, int 16
ITL	buffer for the title, alpha 120
ICL	register of colours for the choropleth maps, int 4
ISY	register of symbols for the choropleth maps, int 4
ICI	store of default symbol types for the choropleth maps, alpha 5

logic

```
request number of maps of real data which are connected
call READ
request elective for graphics backend or end
if end selected jump to end program
call backend & set electives & options for that device
request plot type
request default or user specified plot electives
request number of characters & title
call SWIND
if default electives jump to specified plot call
call INFO
jump to plot type's requests of electives
for each plot type:
     if elective to specify parameters read those requested
     call ISOM or call CONT or call CHOROP
request elective to:
     reselect plot type & jump
     respecify backend & jump
     call READ note file number & last online and jump to
                                                    this menu
end program
request plot of the graphics file if created
```

subroutines

READ

arrays

AR	map matrix for isometrics, real 60,80
AZ	map matrix for contour plots. real 80,60
IAR	buffer for the map array I/O, int 60

logic

```
rewind specified file
read file description
loop
loop read data to IAR
loop
calc element value assign to AR(x,y) and AZ(y,x)
end loops
return
```

ISOM & CONT

arrays

W	workfile for Ginosurf, real 9600
AR or AZ	map matrix arrays, real 60,80 or 80,60

logic

```
call initialising routines
call plot routine
return
```

CHOROP

arrays

IZ	read buffer for maps, int 60
BOUNDS	register of window size, real 4
ITX	store of numeric values & symbols, alpha 7

logic

```
determine size of the plot area by call to SWIND
calc scale factor for the plot to fit current CONT area
(display calc values for match to new devices)
lookup frame colour
move to & draw border of map, move & display label of maps
     from ITL
loop through maps
initialise pen 'x' (NB)
lookup elected pen colour for current device for the current
     map in LC
call SYMBS
rewind and read descriptions of current integer map
loop 'y'
loop read row of map values from current ch to IAR
initialise pen 'y' (NB map rotated in frame)
loop 'x' (NB)
if current value less than or = zero next
if current value greater than six make = six
move pen to current 'y,x'
plot elective symbol at the current point:
     numeric value of the map (lookup in ITX)
     proportional sized selected ASCII symbol
     proportional line shaded box, call SHADE
     proportional sized dot
     proportional stippled box, call STIP
increment 'y'
end loop
increment 'x'
end loop
reset initial choro map plot conditions
end loop
reset to start conditions
return
```

SYMBS

logic

```
position pen for current shade index label
loop symbol sizes
draw current symbol
```

```
increment pen 'x'
end loop
return
```

SHADE

array

ZI	store of 'z' line separations, real 6

logic

```
initialise 'y'
set 'z' increment to current density
draw line
loop
increment 'y' by 'z'
if 'y' is above the height of the symbol box return
reposition 'x'
draw line
end loop
return
```

STIP

arrays

X	store of incremental pen 'x' moves, real 24
Y	store of incremental pen 'y' moves, real 24
MM	store of addresses in X & Y for current move set, int 7

logic

```
lookup current start & stop addresses in MM
move to X(n), Y(n)
invert the sign of the line increment
draw a point line (increment)
increment 'n'
if 'n' = stop value return
jump to move
```

INFO

logic

```
display general information and request plot colour
electively jump to information relative to current plot
display information about electives & formats
request electives
return
```

enhancements

This program is system dependent and its development is not really part of the overall system's future, rather it is to be changed to enable the system to have high definition map output on the devices employed. If possible the facility for pre-stored overlays in the same format as the CHORO program should be implemented.

a3.8 Extract

A program to loop through the entire database picking an elected map of values usually with an associated proximity map. Selection of either real data or real & interpolated data is provided.

arrays

IMAP	map index matrix, int 60,80	
I	output buffer, int 60,12	
K	geological record buffer, int 3	read/write loop implicit
L	header record buffer, int 11	"
LE	engineering records buffer, int 8,5	"
FACT & ITY	map specification registers, real & int 12	
IRG	score register for 'commonest' selection, int 2,10	

logic

```
read selection operation control electives from data file
                                                        ch 10
calc & set special case values encoded in the electives
     e.g. range & specific/general strats. for eng. & lith. data
loop read output map specifications from ch 10
loop write map specifications to output files chs 30-30+n
loop load IMAP & read DA definitions
define DA files
loop 'y'
loop 'x'
loop 'n' and set output buffers to zero
lookup header address in IMAP(x,y)
read header record
check current coordinate and record coordinate if wrong, jump
     to error notes
if serial dump elective write header record to dump file
                                                          ch 30
decode number of eng. & geol. records & pointers
decode top of well & number of boreholes/reliability
if header data elective call HEADR and next
if no layer data or no interpolated elective in operation &
     number of boreholes negative next
if eng. data not required jump to geol. read
loop set eng buffer to zero
loop
read eng. record into strat. specific buffer row
check decoded coordinate matches header, jump to error notes
     if not
if dumping write record to dump file, next
if strat. dependent selection of eng. data call ENGIN, next
end loop
if strat. dependent eng. elective or no real geol. data next
loop
read geol. record
check coord. against header & jump to error notes if not
if dump in operation write record to ch 30, next
```

```
electively call:
     LAYR1 strat. dependent value selection for geol.
     LAYR2 level selected geol. data
     LAYR3 level selected eng. data
     LAYR4 whole well selections of geol data
end loop
if commonest parameter selection not required next
loop through IRG to find largest score and assign the parameter with the highest score to the output buffer
end loop
loop write out 'n' buffers to the appropriate connected map files ch 30-30+n
end loop
jump to end
write out type of error and the actual values of the coordinates & pointers
end program
```

subroutines

HEADR

logic

```
decode elected parameter, with a 'no interpolation' bypass based on the sign if elected, and decode proximity matrix.
(if pH or sulphate and the prox. for a real value is zero, indicating water readings, the prox. map element is set to 100 as an indication)
return
```

ENGIN

logic

```
decode eng. data, with a no interpolated elective bypass, place in output buffer
if worst case elective, select extremes of the averaged parameters not electively excluded from the search for each stored parameter
buffer out the extremes loop selected with their prox. maps
return
```

LAYR1

array

IRS	store of superposition weightings for the strat. codes, int 51

logic

```
set all variables at the start of each new segment
keep a running register of the preceding strat. & layer base level
decode strat. & electively jump, if strat. is elected, to:
     decode & store code & thickness of the layer for lith., strength, colour or specific strat., if code has already occurred in the segment, sum its scores
     pick the first occurrence of the selected strat. in a well & put the value for the base of the preceding strat. in the buffer, set a switch to prevent overwriting of the data
```

pick the lowest occurrance of the strat in a well by setting a switch when the selected strat. occurs and recording the previous layer base when the new strat. is not the same as that required

records the strat. (or custom - lith.) of the layer occurring below (in the predefined sense defined by IRS) that specified (no strat. implication is also written into the current program but unused)

any custom specified (current is lith. below specified strat.)

update the register of layer base height
return

LAYR2

logic

decodes height of base of layer & if the specified level is outside the values of the top & bottom of the layer update the level of the 'top' register and return
update 'top' value
electively decode:

colour, lith., strength or strat. and place in the buffer with the prox. value from the header file for the whole well
return

LAYR3

logic

as LAYR2 but electively decode average & extreme values from the strat in the specified level (calc from the layer data) and put the results into the buffer with the prox value.
return

LAYR4

arrays

LITH	store of grain sizes for the broad litholology, int 10
ISN	store of classified strengths, int 37

logic

set switches if a new segment
electively:

decode the colour, lith., strength or strat. and electively pick out a specific value or one within the range given, write to the buffer with a prox. value from the header distribution

if the decoded strat. of the last layer in the well is that specified, place the level reached in the buffer with a proximity from the header

lookup every decoded lith. (& divide by ten for generic) in LITH, put the worst score in the buffer with prox.

lookup every strength in ISN & place the worst score in the buffer with prox.
return

enhancements

Improved efficiency by the use of a 'no relevant data' bit or existence table/barrier map to limit the wells searched.

A verbal interrogation language for the creation of conditional recall sequences for more complex data combinations. The current level-conditional search modes employ absolute heights O.D.. The conversion of this search to depths below surface is trivial: subtract each level decoded from the current segment surface height and use in the same routines.

a3.9 Extpol And Pxtpol

(Extpol) puts a $\frac{1}{d^2}$ weighted surface through all points within an electively switchable set of ranges from that point and provides a map of zones of proximity of real data to any point in that area. The interpolation may be inhibited in areas indicated by a barrier, or existence, map file.

arrays

IARRAY	master proximity matrix template, int 22,22
LARRAY	three layer array of datapoints, proximity map, barrier map, int 60,80,3
ALL	register of original data values, real 900
IALL	register of coordinates of the data values, real 900,2
DIST	workarea register of separations of real data from the current element, real 900
SEP	register of reliability zone radii, real 6

logic

```
read in electives & number of points to be used in the
     interpolation average
if selected loop in proximity range radii
read in map headers of files selected
write headers to relevant files
calc proximity template constants
loop 'y'
loop 'x'
calc range of current point from centre
assign specific proximity integer to the current element of
     the proximity template
end loops
loop 'y'
loop read points map to LARRAY(x,y,1)
electively loop read barrier map to LARRAY(x,y,2)
loop read data to  LARRAY(x,y,3)
loop 'x'
if a value is present increment score of 'wells'
assign coordinates to IALL(wells,1-2)
calc real value and assign to ALL(wells)
if wells is greater than 900 print note & abandon the run
end loops
if proximity array elective not selected jump to barrier
     process
loop wells
calc main map array address of lower left hand corner of the
     prox. template at the current IALL  x & y coordinates
loop 'y'
if coordinates are outside the main array next
loop 'x'
```

```
if coordinates are outside the main array next
calc current prox. template coords
if prox. value is zero next
if main array value is zero or less than template value set
     the main array to template value
end loops
if no prox. map loop set prox map to -1
if no barrier map elective loop through the prox. & barrier
elements setting prox. to null if barrier map is null (i.e. on)
loop 'y'
loop 'x'
if prox. map is null next
if real data is present in this segment next
loop wells
calc separations of all points to the current element from
     IALL and place in DIST
end loop
loop number of interpolation points to be averaged
pick nearest point
calc weight of point
sum weights
bias array element out of search
end loop
calc total weighted value assign current main data map elem-
     ent
increment number of points interpolated
end loop
loop out data map row
end loop
loop out proximity map
display number of wells & interpolated elements
end program
```

PXTPOL

This is identical to EXTPOL except that once the nearest point is calculated the value is assigned to the output map and the loop continues. The effect obtained is the generation of a Thiessen polygon around each real data point, i.e. the map produced is shaded according to the type of the nearest element to each current point.

enhancements

More interpolation functions, greater sophistication in the selection of the points for averaging. Machine code program for improved speed. Determination of a reliability index for each whole map based on the interpolation factor (NYK), the number of points (NBH or wells) and the mean separation of datapoints (calc by running average from DIST).

a3.10 Replace

Scans through the database re-writing it to a new set of files. This facilitates the re-creation of the datastore DA files from a dump or the streamlining of the datastore by repacking the files in a sequential form after the insertion of a number of records. This reduces the file size and improves the efficiency of serial data searches for maps. The facility for a custom insertion of maps is also provided.

arrays

IMAP	map index matrix, int 60,80
JA	buffer for array maps, if inserting, int 60,10
J	buffer for records, int 14 (I/O loop implicit)
FACT	register for the scales of the maps, real 10

logic

```
specify and read in map definitions if custom
     (NB use the absolute value in encoding and transfer the
     sign last, if map arrays are to be combined in the cust-
     om insertion)
if not reading from serial data read in the map index IMAP
     and the DA parameters
define DA files for input & output
     (NB file pointers are automatically incremented after
     each I/O)
loop 'y'
if custom read in map rows
loop 'x'
look up header address IMAP(x,y)
read in header from serial of DA (serial is dumped on two
     rows of a file by EXTRACT)
check coordinates on the record against the calc N.G. coord-
     inate  if error note and end run
calc number of records & pointers for geol. & eng. files
     (values IRAWL,LAYS & IRAWE,INO)
if no geol. &/ eng. records are present set output header
     elements containing the geol. & eng. record parameters to
     negative values
set map index address to new header pointer
write header to the new header DA file
if no layer data next
if no eng. data jump to geol. read loop
loop INO eng. records
read eng. data from serial or DA, if DA start pointer at IRAWE
check for coordinate match note & end if it occurs
write to new DA file for eng. data
end loop
loop through LAYS geol. records
read geol. data from serial or DA record, at pointer IRAWL
check coordinates match if not note & end
write to the new DA file for geol. data
end loops
loop out new map index
write out DA file maxima to index file
end program
```

enhancements

The simple language or pointer driven insertion of maps into the filestore for replacement & extension, more 'openness'

a3.11 Split

Converts serial coordinated data into map array files.

arrays

AR	register of values for a point, real 3

JAR	output buffer for the three maps, int 60,80,3
ITY & FACT	specification registers for the maps, int, real 3

logic

```
read in number of maps, map specifications
loop
read in National Grid coordinate and 'n' parameters check for
    no more data, if none end
calc array addresses
loop
set current elements to calc integers of parameters
end loops
loop maps
calc ch number
write specification to current file
loop map to current file
end loop
end program
```

enhancements

Accept data from an interactively constructed screen image for data entry or barrier/existence map creation.

a3.12 Read programs

A set of programs for DA file handling, allowing the viewing of these binary files in the same way as the character, or serial, files. They are best held in a compiled form for speed of use.

array

K	buffer for records, int 14 (implicit loop)

logic

```
request the file name, length of records & maximum file
    number for the required file
define the file
either:
    read & display a specified record, as in DREAD
    loop write the whole file to a serial file, as in READ5
    produce an index map (for the header file), as in READ4
end program
```

enhancements

A direct access file character editor.

a3.13 Editor

Reads and displays one full segment on a graphics VDU with an interactive editor provided for alteration and display of alterations to any value in the segment. NB. new records cannot be inserted with this program, SEGMENT should be used for that requirement.

arrays

ISTP	pre-created graphics I/O buffer, int 25
IMAP	store for the map index, int 60,80

logic

```
rewind screen presets & index files
loop read map index & DA parameters
define files
switch on the graphics
loop read user-defined characters, menu etc. from store to
    screen
set background colour & pen colours
request a N.G. coordinate to be retrieved
calc address & reset switches
check coordinate is in the map area & warn & jump to the
    coordinate request if not
lookup header address in IMAP
call READ
jump to request for a coordinate if there is no layer data
    (switch IST)
call ERASE
display current coordinates to screen
call DRAW
call EDIT
depending on the elective from EDIT jump to:
    redraw
    re-read the segment
    new coordinate, jump to N.G. request
    update, call REPLACE then jump to request N.G., set
            screen to black during process
    update & end, as update but jump to end program
    end run
end program
```

subroutines

REPLAC

As WRITE (q.v.) but with a pre-read check of the coordinate for each record replaced. If the value does not match abandon & display a warning. The header data are written to their file after the geol. & eng. data, to fail-safe these file pointers in the event of a geol. or eng. mismatch.

EDIT

arrays

AR	register of geol. layer base heights, real 13
IAR	register of the geol. data codes in the layers of the segment, int 12,9
LE (or LJ)	register of decoded eng. data & general strat key for the segment, int 18,5

logic

```
request cursor input for an elective
if cursor is outside the menu return with a new N.G. elective
calc cursor position to find the elective  selected and jump
All alterations request data & set the specified parameter
register to the new values. See the menu for the electives,
```

most of them are self explanatory. Special processes are:

to relevel all the heights in AR, request a ± X value and loop subtract it from the array values.

Whole record changes, loop read in a set of parameters to the whole register array row.

engineering parameters are read in as real values where relevant.

control electives are transferred to the master routine except the background intensity which calls BACKGR (q.v).

DRAW

logic

```
call WENG
calc top & bottom (for the end correction) heights in the
    well & a scale factor
call SCALE
set number of colours to zero
calc position of movable  labelling
loop
calc:
    position of the top of the layer on the screen
    the height of the layer on the screen
    the position of the centre of the layer
    the number of rows of screen text to cover the layer
                                                   vertically
call DLAY
call WLAY
end loop
display extra header & archive data to the screen
return
```

WLAY

array

LOC store of character widths for spacing, int 9

logic

```
draws a line from DLAY final pen position to calc labelling
    position
loop display a row of codes from IAR, lookup widths in LOC
return
```

WENG

array

LOCE widths store for eng. parameters, int 16

logic

```
moves pen to the calc space
loop display labels for the header stored general parameters
move pen to a new line
display header general eng. data
move pen to a new line
loop display a label for the eng. data
move to a new line
loop
display a row of LJ to the screen
end loop
return
```

enhancements

Display realism and flexibility. Allow replacement of data by averaging-in. Show more parameters schematically, after Bouma & Nota (1961).

a3.14 Merge

A data preparation program, allowing the synthesis of one set of lithological records from two or more borehole logs using a computer-aided 'pattern recognition' feature selection. Up to nine boreholes with eleven layers each may be displayed & processed at one time to give up to 24 layers out.

arrays

ARR	register of levels of layers in the input boreholes, real 9,12
IARR	register of coded information for the input boreholes, int 9,11,21
AR	register of levels in the synthesised well, real 25
IAR	register of geol. information for the synth. well, int 24,9
IHE	register of the number of layers and the borehole reliability index for the input boreholes, int 9,2
ISAR	register of the layers in the well which the input borehole's layers have been assigned to, int 9,11
IHZ	store of weightings for the types of made ground, int 21
ISTP	buffer for the graphics setup, int 25

logic

```
read in the coordinates & pointers from the dump file
loop position pointers in the in & output files.
buffer through the graphics preset values & background
calc number of segments left to process
loop remaining segments
read in header data
if no layer data jump to header write then next
loop in eng. data for each borehole into IARR
loop in geol. data for each borehole into IARR & ARR
pick out the worst case header data (made type, worst levels
     the pH, SPT, sulphate have already been processed)
allow an interactive synthesis if there are more than one
     boreholes by calling SCAN
write out header values
loop out eng. data for each borehole plus assignments (ISAR)
loop out geol. data for the well (AR & IAR)
update the dump pointer file
end loop
end program
```

subroutines

SCAN

arrays

IRG	register of parameters & scores for the commonest parameter selection, int 8,9,2
NH	register of number of different parameters, int 9

logic

```
calc scale factors & end correction for the display
loop
increment displayed borehole 'x'
loop
calc 'y' position of the current layer and the thickness in
    real & character terms
call WLAY
end loops
loop synthetic well layers
warn of potential overflow in the number of layers
request cursor selection of borehole layer or a menu item
electively:
    calc borehole & layer hit, then: mark as hit, display
            parameters of the layer, register assignment
            in ISAR(layer, hole), loop through parameters
            scoring & summing scores for each parameter
            and type, warn of potential overflows in the
            parameter score registers, place a pointer
            at the borehole depth 'used'.
    loop select the commonest recorded type recorded for
            each parameter & put in IAR for the current
            layer, calc the average of the lowest levels
            reached in all the boreholes and assign to
            AR, display the new layer (call WLAY).
    request a new position for the base of a layer in the
            synth. well, place in AR & redraw the layer
    request parameters to redefine the values for the curr-
            ent layer & place in IAR.
    display eng. data cross reference file ISAR and request
            the addresses to be altered.
    display a full listing of the data codes for a requested
            layer.
    allow a restart of the synthesis, jump to the beginning
            of the synthesis.
    indicate that the last hit missed the menu & holes,
            request a hole & layer and jump to the layer
            hit routine.
    crash stop, return to the master program and jump to
            the end without writing the segment or updat-
            ing the pointer file.
    normal stop, returns to the master, allows write-out and
            then terminates.
```

The write-out, return for a continuation is automatic in the program.

enhancements

The program should be reconstructed to enable it to use variable input formats. The averaging of engineering data and the check program's (AVER) facilities could be added to this program and the whole package used as an interactive insertion program to allow site investigation reports to be read directly into the database.

a3.15 Aver

A program to check the records output by MERGE and the single borehole segments for errors via an automated human visual processing. Alteration of the parameters to correct or improve the synthesised well is provided for. The eng. data are averaged (although this information was later replaced by AVERAG, q.v.) to provide a total worst case average for each parameter. The results from this program are written directly to the DA datastore.

arrays

as MERGE for ARR,AR,IARR,IAR,IHE,ISAR,ISTP

IER	buffer for the 'previous' pointers for crash dumping, int 7,2
ICHK	character widths for all the parameters in the DA files, int 21

logic

```
define DA files
read pointers from dump file
loop position the serial file pointers
buffer through the graphics preset values & background
calc number of segments left to be processed
loop through the remaining segments
set switches
read in header
set dump output register
if no wells jump to header write out then next
read in the rest of the header data, as derived from the
     wells
loop read in the eng. data
loop read in the geol. data
(check & auto correct for level errors due to the previous
     processor in single wells)
read in original borehole data if there is more than one
     borehole (i.e. synthesis)
initiate interactive check if more than one hole, call SCAN
update dump file & crash if the elective is returned to do so
call AVERAGE
make up the well-derived header buffer
check for widths liable to cause contamination during packing
     (ICHK lookup & warn)
write out the eng. data record
loop encode and write out geol. records
write out the header records  (the format has subsequently
     been altered)
update the dump file
end loop
end program
```

subroutine

SCAN

array

ARB secondary register for the well layer base levels, real 25

logic

```
loop find top & bottom of all the boreholes
calc end correction & scale factor for the screen display
check for errors in the eng. data assignment (errors in ISAR)
loop
increment borehole drawing location
loop
calc height, centre & character height of the current hole
call WLAY
end loops
erase the previously displayed well
loop calc & draw the current synth. well
if errors in ISAR, call ASSCHK (allows reassignment of layers
     & holes in ISAR at the user's discretion)
request cursor input, call CURSOR
     read coordinates hit, calc position on screen, if menu
     area calc which and return the significant value, else
     return significant integers for other areas
electively (depending on the significant integer):
     indicate the layer which has been hit, list all geol.
               data for that area
     request a new level for the base of the hit layer
     loop draw a new synth. well (at the r.h.s. of the screen)
     call ASSCHK
     relevel the whole well by the specified increment
     list original & request a new whole layer record
     respecify the colours for one layer
     respecify the stratigraphy for one layer
     register no hit on a layer & request that aimed at
     normal end of run (write-out current, update pointers)
     output current well, next
     crash the program, no update or data write-out
return
```

enhancements - see MERGE

a3.16 Averag

This program averages the engineering data for a segment from 1 to 9 original boreholes. It uses a list of well stratigraphies to divide the data up into separate sets of values for each strat. type. For each set of data the average and extreme values are calculated. The results are coded into the the same format as the DA eng. file for direct insertion via REPLACE.

arrays

IARR,IAR,ISAR,IHE as MERGE except the no geol. data is held in IARR & IAR, reducing their dimensions to 9,13,12 & 12,12. IHE contains no reliability and is therefore monodimensional & nine characters long.

IOUT	output buffer for the eng. results, int 12
IFIN	matrix of reduced values for selection of the worst cases, int 4,4
IST	register of stratigraphy & layers in that strat., int 6,2
IC	factor store for correction of the biases on the BD & Cu, int 4

logic

```
loop
initialise
read in eng. header from the eng. file
read in strat. information & loop layers
check that the coordinates match, if not warn & abandon
loop in the eng. data (NB bulk density (BD) and shear stren-
     gth (Cu) are given as negative values)
loop check for the presence of any data, if none next
loop set synth. well register to zero
loop boreholes
loop layers
loop parameters
if 'extreme' parameter value next
lookup relevant IAR addresses in ISAR & sum values from the
     current element of IARR into IAR
end loop
select extreme values from relevant columns
end loops
loop synth. layers
loop average column
calc averages
end loop
preset current start layer to 1
loop synth. layers
set current stop layer from IST(n,2)
loop set output matrix to zero
bias BD & Cu elements of matrix to $-1X10^{6}$
initialise registers
loop pick lowest averages ≠ zero & set current row in output
     matrix to selected values
calc worst Atterberg values & set matrix (IFIN) row
end loop
update current start layer
if no data next
reset BD & Cu bias if unused
loop across output matrix
code current matrix elements into the current & current+4
     output buffer with lookup (IC) correction for negative
     bias to BD & Cu
write out coodinate, stratigraphy & buffer to eng. output
     file
increment number of records
end loop
output number of records to the register file
loop set register IST to zero
end loop
end program
```

enhancements - see MERGE

a3.17 Common Subroutines

a3.17.1 Backgr

logic

```
request elective to increment, decrement or return
check that calc new RGB will not be out of range
increment colour intensity registers
set colours on screen
jump to request
```

a3.17.2 Erase

logic

```
draw a block of defined size in the background colour to
     obliterate the previous image, its size depending on
     predefined and supplied screen coordinates
return
```

a3.17.3 Dlay / wlay / llay

logic

```
if there is a primary colour call COLOUR
call LITH
display current strat. code to one side of the centre or top
     of the displayed layer
if a second colour is present call COLOUR and display a
     coloured character sized or equal to the current qual-
     ifier
number the layer in its top right hand corner
position pen at start of pointer line if in EDITOR
return
```

a3.17.4 Colour

arrays

IPC	register of colour RGB and WPR'72 numeric equivalent, int 15,4
ICL	current RGB buffer, int 3

logic

```
if no colour given return
loop through colour register
if value has already been specified in the current well or
     loop return with the VLT address (or colour code)
increment the VLT address register
unpack the light/dark, primary and secondary elements of the
     equivalent code
call MUNSEL
if number of colours is at the maximum display note
loop set the RGB and equivalent in the register
set the current VLT address to the calc RGB
return the new colour address, return
```

a3.17.5 Munsel

arrays

IC	store of RGB codes subjectively determined for the WPR'72 colours, int 3,10,2
I	current colour register, int 2,3
ICL	current RGB, int 3

logic

```
if dark set lightness integer to -2
if orange is specified in the subordinate colour swap prim-
    ary & secondary colours
loop RGB values for primary colour into register (n,1)
if no subordinate colour jump to light/dark set
loop
put RGB for second colour into register (n,2)
average successive R,G,B values and add the lightness integer
    into ICL
end loop
jump to check
loop add lightness to the primary colour only & put in ICL
loop check for ICL values below zero or too high & assign
    out of range values to zero or 15 (max.)
return
```

a3.17.6 Lith

arrays

IL	store of type proportions to the nearest 25% in lithology textures, int 4,95
IT	store of character equivalents for the types on the graphics VDU (as specified in the graphics preset), alpha 12
IT2	store of LP character types, alpha 12
ILOP	store of permutations of texture type character order, int 4,12
IBF	encode/decode buffer for the FORTRAN display format transfer, int 12
ILI	buffer for the made-up character texture string, alpha 4

logic

```
if the value specified is zero set to a blank texture type
set background colour to specified value
loop required number of rows
sequentially place custom symbols for contents type (from IT
    or IT2) of deposit into an output array in proportion &
    type specified in ILI and repeat sequence from ILOP
      i.e. IT(IL(ILOP(broad count, fast count),lithology))
decrement display position
display made-up row to current device
end loop
return
```

a3.17.7 Read

arrays

AR,IAR,LE,J as described in SEGMENT or EDITOR

logic

```
read header file record specified to J
check coordinates in record vs specified & warn if mismatch
decode header parameters into real & int variables
if no layer data set switch IST and return
decode pointers & numbers of eng. & geol. records
if no eng. data jump to geol. read
loop
read eng. record into J
decode coordinate
check against header & warn of mismatch
loop
decode average values & numbers of values to LE
decode extreme value & layer (if real data) to LE
end loops
loop
read geol. record into J
check coordinates & warn of mismatch
decode elements of layer data into IAR & AR
end loop
return
```

a3.17.8 Write / replac

arrays as READ

logic

NB all codes use the absolute value of all parameters and transfer the sign indicating the interpolation as the last step.

```
code all header data values into output buffer J
write out record to header file
if no geological layers present return
loop
code strat. and coordinate
loop code average & number from LE
loop code extreme & layer from LE
write out record to eng. file
end loop
loop
code layer data from IAR & AR
write out record to geol. file
end loop
return
```

The DA pointers are placed at the end of the files for WRITE in the SEGMENT program and their final values made the maxima for subsequent runs. In EDITOR the existing records are read by the routine and the coordinates checked against the current segment being edited before writing takes place (NB backspace pointers after the read) the header is written last so that if an error occurs in the eng. or geol. write loops the pointer values in the header file are not corrupted.

a3.17.9 Scale

array

Optional decode register for LP writing of integers in an alphanumeric form.

logic

```
integerise the level of the top of the scale
make real again
if real height is less than the base of the scale return
calc screen or buffer coordinates from the real value
move to coordinate & write real value (split up int. value
    into elements & loop into current display buffer row)
subtract scale interval from the integer height
jump to 'make real'
```

To mark off points on the scale calc scaled & corrected 'y' and move to appropriate, display centred symbol.

a4 FICHE LISTING OF THE PROGRAMS

Contents:

MAIN PROGRAMS

name	length
mapinfo	275
codeinfo	193
proginfo	201
startup	99
combine	220
rcombine	20
ecombine	33
scatter	155
rscatter	10
choro	173
rchoro	14
eg VLT	16
section	606
rsection	12
tsection	90
segment	428
rsegment	6
tsegment	192
draw	413
rdraw	42

BATCH PROGRAMS

name	length
extract	515
rextract	11
pextract	256
dextract	18
bextract	22
extpol	139
rextpol	11
pextpol	53
dextpol	16
bextpol	22
pxtpol	128
rpxtpol	11
bpxtpol	22
replace	85
rreplace	23
bseplace	12
bbeplace	15

SERVICE PROGRAMS

name	length
three	213
rthree	13
read4	23
read5	17
rread5	16
dread	6
read	21
coltab	28
rcoltab	11
col	110
rcol	8
split	20
munsell	84
rmunsell	8
editor	733
reditor	25
teditor	178
psout	10
rpsout	8
screens	21
lamb	81
rlamb	12

DATA REDUCTION

mktrans
fptrans
read (part)
merge
bit (tmerge)
aver
bit2 (taver)
averag

PLOTTING PROGRAMS

mdif
dplot
hdplot
fastplot
tile

'length' is in lines

EAEXY
DVLIST

a5 FICHE LISTING OF CHORO & SCATTER PLOTS

EAEXX
DVLIST

Contents:

519 lines of CHORO maps containing the three surfaces which make up the three dimensional model of the 'map' and the 'reality' maps for the three stratigraphies. The maps are in the order: Made, Recent, Glacial and are divided into fifteen classes.

2254 lines of CHORO maps of the variables in the hazard map with their associated proximity maps. The maps are in the order: Made isopach, cover isopach, sulphate, pH, archive, rock type, SPT, shear strength, qualitative strength, consistency index, plasticity index, liquid limit, saturation, unit weight. All the maps are divided into five classes and are of the worst case data.

290 lines of SCATTER plots for the: Made isopach, cover isopach, pH, SPT, shear strength, consistency index, plasticity index, liquid limit, saturation, unit weight, moisture content.

865 lines of CHORO plots each classified into ten classes which show the build up of the hazard map and its proximity map. Also, 406 lines of SCATTER plots from each stage of the summing of the hazard map.

579 lines of SCATTER and five-class CHORO maps of the various results from the weighting factors tested before the final fourth power skew was found. Each map has a SCATTER plot and a proximity map.

756 lines of proximity maps for the made, recent and glacial isopachyte and deposit type maps used in the composite maps in Appendix 7.

a6 FICHE LISTINGS OF THE DATA

a6.1 Original data in uninterpolated form. The format may be found in Appendix 2. The Header data is split over two lines.

EAEXX DVLIST

a6.2 Interpolated database listing. The format may be found in Appendix 2. This listing is on three sheets.

EAEXX
DVISTS

REFERENCES

Ammer, Becht, Bents, Rosenkrantz & Rosenkrantz (1976): Ecological mapping of the community: Elaboration of a scheme for the classification of community territory on the basis of its environmental characteristics. Brussels Comm. EEC.

Bazley R.A.B. (1971): A map of Belfast for the engineering geologist. Q. Jl. Engng. Geol., v.4, 313-314.

Bennett D.H. (1979): Site investigation report for the Ladywood valley sewer. Strata Test Co. Ltd. pp 12.

Berk T., Brownston L., Kaufman A. (1982): A new colour-naming system for graphics languages. IEEE Computer Graphics & Applications, v.2(3), 37-44.

Bouma A.H. & Nota D.J.G. (1961): Detailed graphic log of sedimentary formations. 21st. Int. Geol. Cong. Copenhagen, pt.23, 52-74.

Bracinac Z. & Janjic M. (1978): Engineering geological maps of seismic regions. Bull. Int. Assoc. Engng. Geol. No. 18, 27-32.

Brisbin W.C. & Ediger N.M. (1967): A national system for the storage and retrieval of geological data in Canada: report of the Ad Hoc Committee on storage and retrieval of geological data in Canada. Geol. Surv. Can., pp 175.

British Standard Code of Practice, CP2001 (1957): BS. 5930 Site Investigations. British Standards Institute, London.

Brunsden D., Doornkamp J.C., Fookes P.G., Jones D.C.K., Kelly J.M.H. (1975): Large-scale geomorphological mapping and highway engineering design. Q. Jl. Engng. Geol., v.8, 227-253.

-----------. (1975): Geomorphological mapping techniques in highway engineering. The Highway Engineer, v.22, No.12, 35-41.

Buisson J.L., Gros G., Sanejouand R., & Voiment R. (1979): Computer-aided updating of the engineering geological map of Rouen. Bull. Int. Assoc. Engng. Geol., No.19, 303-311.

Buller J.B. (1967): A computer orientated system for the storage and retrieval of well information. Bull. Can. Petrol. Geol., v.12, 847-891.

Cantell T. (1977): Urban wasteland: a report on land lying dormant in cities, towns and villages in Britain. Civic Trust, pp 56.

Chadwick P.K. (1975): Psychological analysis of observation in geology. Nature, London, No.256, 570-573.

---------. (1976): Psychological analysis of illusions in geology. Nature, London, No.260, 397-401.

Chandler R.J., Birch N. & Davis A.G. (1968): Engineering properties of the Keuper Marl. CIRIA research report, No.13.

Coe L., & Cratchley C.R. (1979): The influence of data point distribution on automatic contouring. Bull. Int. Assoc. Engng. Geol., No.19, 284-290.

Computer-Aided Design Centre (1976): GINO-F user manual. CADCentre, Cambridge, pp 150.

---------. (1981): GINOSURF user manual. CADCentre, Cambridge, pp 82.

Cottiss G.L., Dowell R.W. & Franklin J.A. (1971): A rock classification system applied in civil engineering. Civil Engineering & Public Works Review, v.66, 611-614 & 736-743.

Cratchley C.R., Conway B.W., Northmore K.J. & Denness B. (1979): Regional geological and geotechnical survey of south Essex. Bull. Int. Assoc. Engng. Geol., No.19, 30-40.

Cripps J.C. (1979): Computer storage of geotechnical data for use during urban development. Bull. Int. Assoc. Engng. Geol., No.19, 290-295.

Cutbill J.L. (ed.) (1971): Data processing in biology and geology. Systematics Association Spec. Vol. No.3, Academic Press, London, pp 346.

Davis J.C. (1973): Statistics and data analysis in geology. Wiley, New York, pp 550.

Davis J.C. & McCullagh (eds.) (1975): Display and analysis of spatial data. Wiley, London, pp 378.

Dearman W.R., Money M.S., Coffey J.R., Scott P. & Wheeler M. (1973): Techniques of engineering geological mapping and examples from Tyneside in The engineering geology of reclamation and development. Reg. Mtng. Durham Engng. Gp. Geol. Soc. London. 31-34.

Dearman W.R. & Matula M. (1976): Environmental aspects of engineering geological mapping. Bull. Int. Assoc. Engng. Geol., No.14, 141-146.

Dearman W.R, Money M.S., Coffey J.R., Scott P.J. & Wheeler M. (1977): Engineering geological mapping of the Tyne and Wear conurbation, N.E.England. Q. Jl. Engng. Geol., v.10A, 145-168.

Dearman W.R., Money M.S., Strachan A.D., Coffey J.R. & Marsden A. (1979): A regional engineering geological map of the Tyne and Wear county, N.E.England. Bull. Int. Assoc. Engng. Geol., No.19, 5-18.

Dillon E.L. (1967): Expanding role of computers in geology. Am. Ass. Petrol. Geol. Bull., v.51, 1185-1201.

Duigan S.L. (1956): Pollen analysis of the Nechells interglacial deposits, Birmingham. Q. Jl. Geol. Soc. London, v.112, 373-391.

Eastwood T., Whitehead T.H. & Robertson T. (1925): The geology of the country around Birmingham. Mem. Geol. Surv. G.B., pp 152.

Experimental Cartography Unit (1969): Automatic cartography and planning. Roy. Coll. Art, Architectural Press London, pp 150.

Fisher H.T. (1968): Reference manual for "SYMAP" version 5. Lab. for Comp. Graphs. Sch. Design Harvard Univ., pp 135.

Fookes P.G. (1967): Planning and stages of site investigation. Engng. Geol. v.2, 81-106.

Franklin J.A., Broch E. & Walton G. (1971): Logging the mechanical character of rock. Trans. Inst. Min. Metall., v.80, A1-9.

Freeman H. & Peroni G.G. (eds.) (1979): Map data processing. Proc. N.A.T.O. Adv. Study Inst. Maratea, Italy, Academic Press, New York, pp 374.

Geological Society of America (1970): Rock colour chart, Geol. Soc. Am., Boulder, Colorado.

Glossop R. (1969): Engineering geology and soil mechanics. Q. Jl. Engng. Geol., v.2, 1-5.

Gounon A. (1979): Taking account of natural hazards for urban planning documents: an example from the city of Nice. Bull. Int. Assoc. Engng. Geol., No.19, 126-128.

Gover T.N., Read W.A. & Rowson A.G. (1971): A pilot project on the storage and retrieval of geological information from cored boreholes in Central Scotland. Inst. Geol. Sci. Rept. 71/13.

Griffiths J.C. & Rosenfeld M.A. (1954): Operator varation in experimental research. Jl. Geol., v.62, 74-91.

Harbaugh J.W. & Merriam D.F. (eds.) (1968): Computer applications in Stratigraphic analysis. Wiley, London, pp 282.

Harrison W.J. (1894): Bibliography of Midland glaciology, Proc. Birm. Nat. Hist. & Phil. Soc., v.ix, 116-200

Harrison W.J. (1898): Ancient glaciers of the Midland counties of England. Proc. Geol. Ass. London, v.15, 400-408.

Horton A. (1974): The sequence of the Pleistocene deposits proved during the construction of the Birmingham Motorways. Inst. Geol. Sci. Rept. 74/11.

Horton A. (1975): The engineering geology of the Pleistocene deposits of the Birmingham district. Inst. Geol. Sci. Rept. 75/4.

Hubaux A. (1973): A new geological tool - the data. Earth Science Review, v.9, 159-203.

Hutchinson J.N. (1977): Assessment of the effectiveness of corrective measures in relation to geological conditions and types of slope movement. Bull. Int. Assoc. Engng. Geol., No.16, 131-155.

Hwong T.C. (1978): Classification of the rock mass structures and a determination of rock mass quality. Bull. Int. Assoc. Engng. Geol., No.18, 139-142.

International Business Machines (1965): Numerical surface techniques and contour mapping. IBM D.P. Appls. White Plains, New York.

Jeffery G.K. & Gill E.M. (1977): The design philosophy of the G-EXEC systen. Computers & Geosciences, No.2, 345-346.

Kelly M.R. (1964): The middle Pleistocene of North Birmingham. Philos. Trans. Roy. Soc. London, v.B247, 533-592.

Krumbein W.C. (1959): Trend surface analysis of contour-type maps with irregular control-point spacing. Jl. Geoph. Res., v.64, 823-834.

Lopez Prado J. & Pena Pinto J.L. (1979): Problems involved in the preparation of geological maps at a scale of 1:25,000. Bull. Int. Assoc. Engng. Geol., No.19, 84-87.

Loudon T.V. (1976): The ROCKDOC package. Reading Univ. Geol. Rept. No.1.

---------. (1979): Computer methods in geology. Academic Press, London, pp 269.

Martin D.H. (1980): Database design and implementation on maxi and mini computers. Van Nostrand Reinhold, London, pp 144.

Matula M. (1979): Engineering geological evaluation for regional and urban development. Bull. Int. Assoc. Engng. Geol., No.19, 1-5.

Merriam D.F. (ed.) (1969): Computer Applications in the Earth Sciences. Plenium Press, London & New York, pp 281.

Morin F.J. (1979): Computerized automatic geotechnical mapping from a geoscientific databank. Bull. Int. Assoc. Engng. Geol., No.19, 319-322.

Munsell A.H. (1954): A colour notation. Munsell soil colour charts, Munsell Color Co. Ltd., Baltimore.

Newman W.M. & Sproull R.F. (1979): Principles of interactive computer graphics. 2nd. ed., McGraw Hill, London, pp 541.

Odell J. (1977): Description in the geological sciences and the Lithostratigraphical descriptive system LSD02. Geol. Mag., v.114, 81-114.

Oxford University Press (1963): Atlas of Britain.

Pickering R. (1957): Pleistocene geology of the south Birmingham area. Q. Jl. Geol. Soc., v.113, 223-237.

Pocock R.L. (1979): The measurement and prediction of ambient environmental conditions in urban areas. unpub. Ph.D. thesis, Univ. of Aston, pp 542.

Price D.G. (1971): Engineering geology in the urban environment. Q. Jl. Engng. Geol., v.4, 191-208.

Reekie C.J., Coffey J.R. & Marsden A.E. (1979): Computer-aided techniques in urban engineering geological mapping. Bull. Int. Assoc. Engng. Geol., No.19, 322-330.

Rhind D.W. & Sissons J.B. (1971): Data banking of drift borehole records for the Edinburgh Area. Inst. Geol. Sci. Rept. 71/15.

Rockaway J.D. (1976): Influence of map scale on engineering geological mapping. Bull. Int. Assoc. Engng. Geol., No.14, 119-122.

Rosing K.E. & Wood P.A. (eds.) (1971): Character of a conurbation. in A computer atlas of Birmingham and the Black Country. Univ. of London Press, London, pp 126.

Salford Computing Centre (n.a.): GINOSURF manual. repub. as Computer-aided design centre GINOSURF users manual, 1981.

Shepard D. (1968): A two dimensional interpolation function for irregularly-spaced data. Proc. 23rd. Nat. Conf. Ass. Comp. Mach., 517-524.

Shotton F.W. (1953): The Pleistocene deposits of the area between Coventry, Rugby and Leamington and their bearing upon the topographic development of the Midlands. Phil. Trans. Roy. Soc., v.B237, 209-260.

Sokal R.R. & Rohlf F.J. (1970): The intelligent ignoramus an experiment in numerical taxonomy. Taxon., v.19, 305-319.

Sowerbutts W.T.C. & Mason R.W.I. (1981): A mobile microcomputer based system for on-site recording, processing and plotting of small-scale geophysical survey results. Comp. Appls. in Geol. III, AGM. Geol. Soc. London, Geol. Info. Gp.

Stewart R.M., Hart E.W & Amimoto P.Y. (1976): The review process and the adequacy of geologic reports. Bull. Int. Assoc. Engng. Geol., No.14, 83-88.

Taylor P.J., Kirby, Harrop & Gudgin (eds.) (1976): Atlas of Tyne and Wear, Newcastle-upon-Tyne dept. of Geog., pp 37.

Tuckwell D.J. & Sadegrove B.M (1977): A case for a National Registry of Ground Investigation Reports. CIRIA rept. No.70, pp 21.

UNESCO/IAEG (1976): Engineering geological maps. A guide to their preparation. The UNESCO Press, Paris, pp 79.

Varey N. (1977): A computerised storage and retrieval system for the communication of geotechnical data. unpub. Ph.D. thesis, Univ. of Aston, pp 287.

Vidal-Font J. (1979): A tentative geotechnical valorisation of the geologic map of Toulouse (France) and the surrounding area. Bull. Int. Assoc. Engng. Geol., No.19, 53-57.

Ward W.H., Burland J.B. & Gallois R.W. (1968): Geotechnical assessment of a site at Munford, Norfolk, for a large proton accelerator. Geotechnique, v.18, 399-431.

Wills L.J. (1937): The Pleistocene history of the west Midlands. Rept. Brit. Assoc. Adv. Sci., Notts., 71-94.

Wilson H.E. (1972): The geological map and the engineer. Proc. 24th. Int. Geol. Cong. Montreal, sect.13, 83-86.

WPR'70 or Anon. (1970): The logging of rock cores for engineering purposes. Q. Jl. Engng. Geol., v.3, 1-24.

WPR'72 or Anon. (1972): The preparation of maps and plans in terms of engineering geology. Q. Jl. Engng. Geol., v.5, 293-381.

Zetter J.A. (1974): The application of potential surface analysis to rural planning. The Planner, v.16(2), 544-549.

BIBLIOGRAPHY

Aston Computer Centre (annual): A general introduction to the computing facilities at Aston. Aston Comp. Service.

Aston Computer Centre (1976): The GEORGE 3 operating system. Aston Comp. Service.

Aston Computer Centre (1978): FORTRAN. Aston Comp. Service.

Birmingham City Council: County Structure Plan Review.

Burnett A.D. & Fookes P.G. (1974): A regional engineering geological study of the London Clay in the London and Hampshire basins. Q. Jl. Engng. Geol., v.7,257-295.

CalComp Ltd. (1971): A GPCP user's manual. CalComp Co. House.

Christiansen E.A. (ed.) (1970): Physical environment of Saskatoon. N.R.C. pub. No. 11378, Ottawa-Canada, pp 68.

Cratchley C.R. (1977): Engineering geology of South Essex in Inst. Civ. Eng., Surface Modelling by Computer. I.C.E. Paper No. 6, 43-49. and Discussion, 92-99, Inst. Civ. Eng. London.

Dearman W.R. (1974): Weathering classification in the characterisation of rocks for engineering purposes in British practice. Bull. Int. Assoc. Engng. Geol., No.9, 33-42.

De Moor G. & De Breuck W. (1976): Preparation of semi-detailed lithologic and hydrogeologic maps for land management by means of a geo-electrical survey. Bull. Int. Assoc. Engng, Geol., No.14, 137-140.

E.E.C. (1976), The South Yorkshire environmental study.

Fookes P.G., Hinch L.W. & Dixon J.C. (1972): Geotechnical considerations of site investigation for stage IV of the Taff Vale trunk road, in South Wales. 2nd. Brit. Reg. Cong., Perm. Int. Ass. Road Congs. Cardiff, 1-25.

Franklin J.A. (1970): Observations and tests for engineering description and mapping of rocks. Proc. 2nd. Int. Cong. Rock Mech. Belgrade, paper 1-3, 1-6.

Grant F (1957): A problem in the analysis of geophysical data. Geophysics, v.22, 309-344.

Harris Computer Systems (1980): FORTRAN 66 reference manual. Aston Comp. Services.

Harris Computer Systems (1980): Vulcan job control reference manual. Aston Comp. Services.

International Computers Limited (1971): Introduction to 1900 Extended FORTRAN. I.C.L. Technical Pub. 4269, pp 92.

Leith C.J., Schneider K.A. & Carr C. (1976): Geophysical investigation of archaeological sites. Bull. Int. Assoc. Engng. Geol., No.14, 123-128.

Loudon T.V. (1974): Analysis of geological data using ROCDOC, a FORTRAN IV package for the IBM 360/65. Inst. Geol. Sci. Rept. 74/1.

McLain D.H. (1974): Drawing contours from arbitrary data points. Computer Journal, v.17, 318-324.

Merla A., Merlo C. & Oliveri F. (1976): Detailed engineering-geological mapping in selected Italian mountainous areas: methodology and examples. Bull. Int. Assoc. Engng. Geol., No.14, 129-135.

Planning and Transport Research and Computation Co. (1973): Surface and sub-surface survey and mapping. summer annual mtng. Univ. Sussex, PTRC Co.

Rodriguez Ortiz J.M. & Prieto C. (1979): A proposal for quantitative terrain eveluation and highway construction. Q. Jl. Engng. Geol., v.12, 139-146.

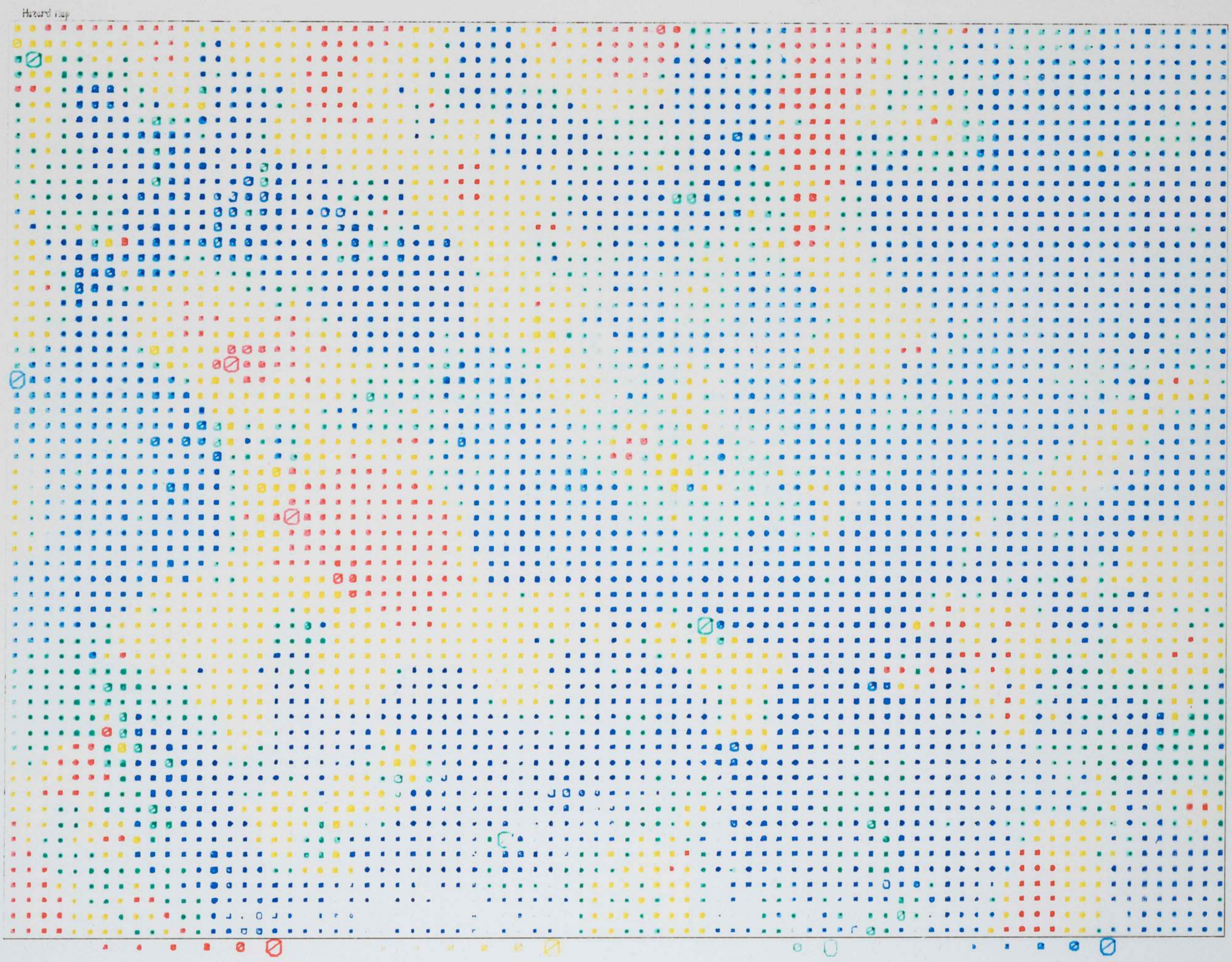
Sylvester-Bradley P.C. & Ford T.D. (eds.) (1968): The geology of the East Midlands. Leicester Univ. Press, pp 400.

Walters R.F. (1969): Contouring by machine (a user's guide). Am. Ass. Petrol. Geol. Bull., v.53, 2324-2340.

Warner J.R. (1981): Principles of device independent computer graphics software. IEEE Computer Graphics & Applications, v.1(4), 5-18.

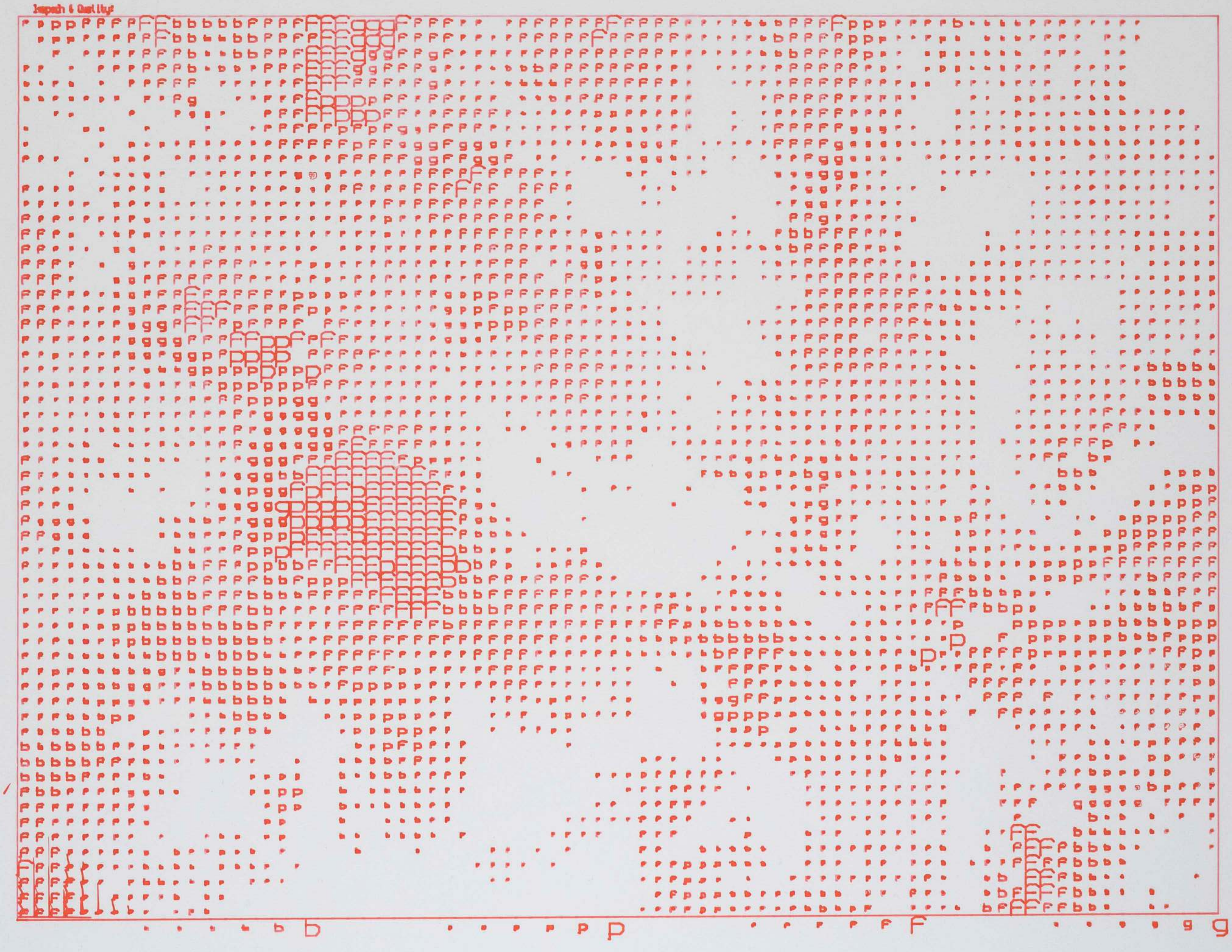
Waters R.S. (1958): Morphological mapping. Geography, v.43, 10-17.

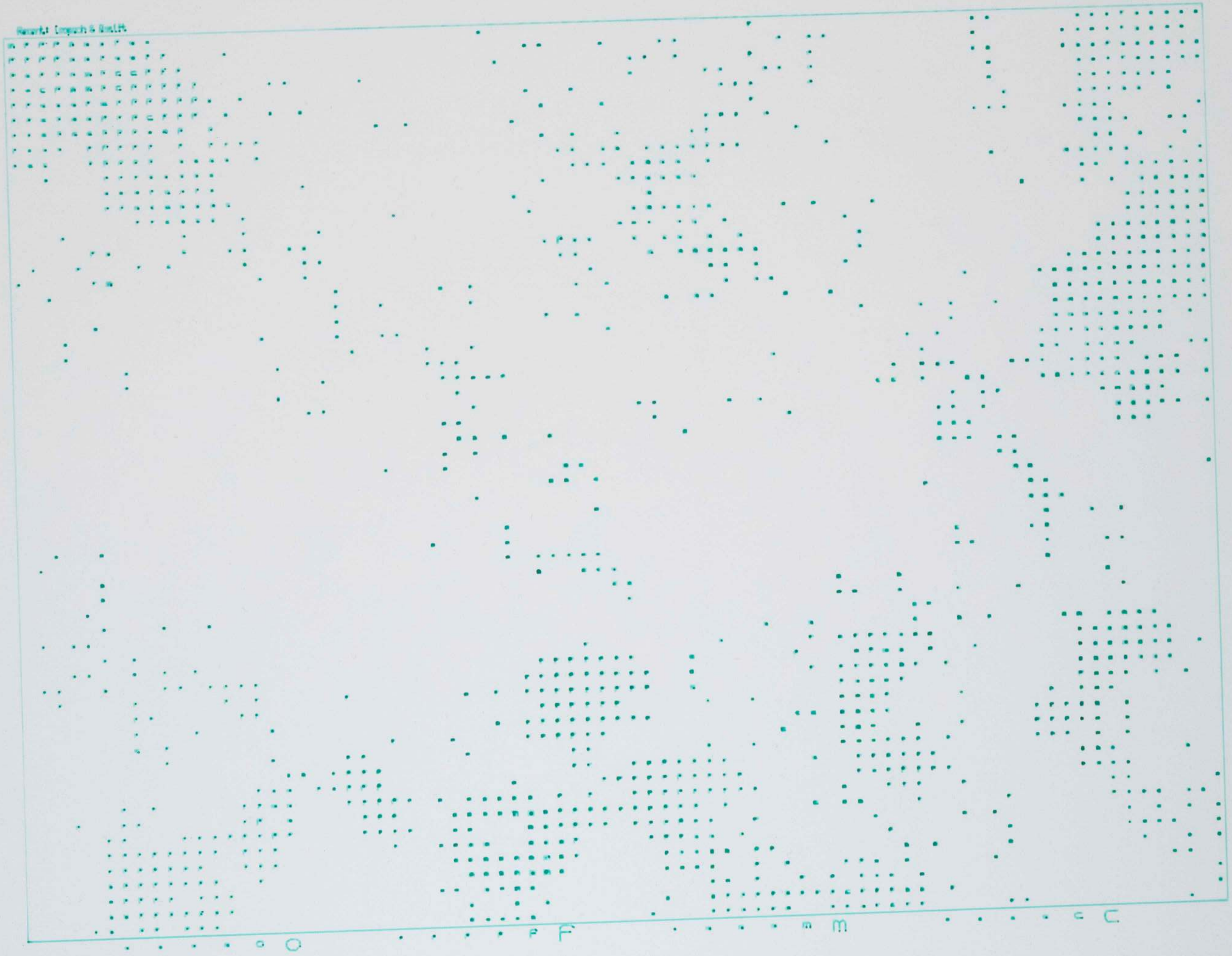
Wills L.J. (1950): The Palaeogeography of the Midlands. Univ. Press, Liverpool, pp 147.


Hazard map


110
110
110
110
110
120
110
110
110
110
100
100
100
110
100
100
100
90
90
90
100
90
90
90
100
90
90
90
100
100
90
90
90
90
90
100
110
100
100
110
110
100
90


120
120
120
110
130
120
120
110
130
110
120
120
110
120
20
110
120
130
110
110
130
130
120
130
110
120
110
120
110
110
120
100
100
130
110
110
100
110
120
100
110
100
100
100
110
120
120
90
100
100
90
120
90
100
100
110
90
100
110
120
120
110
100
100
110
110
120
100
100
120

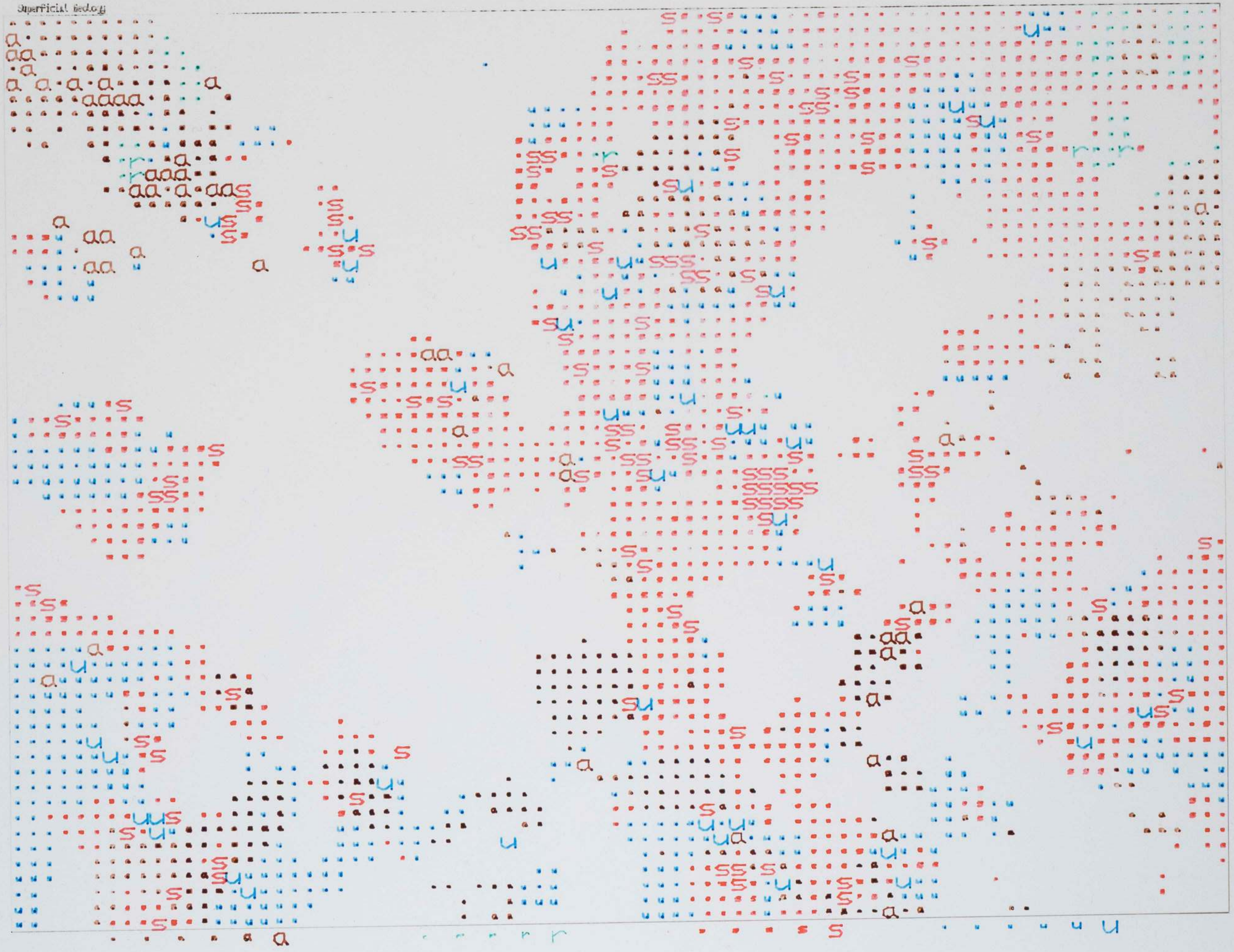

Impach & Quality?

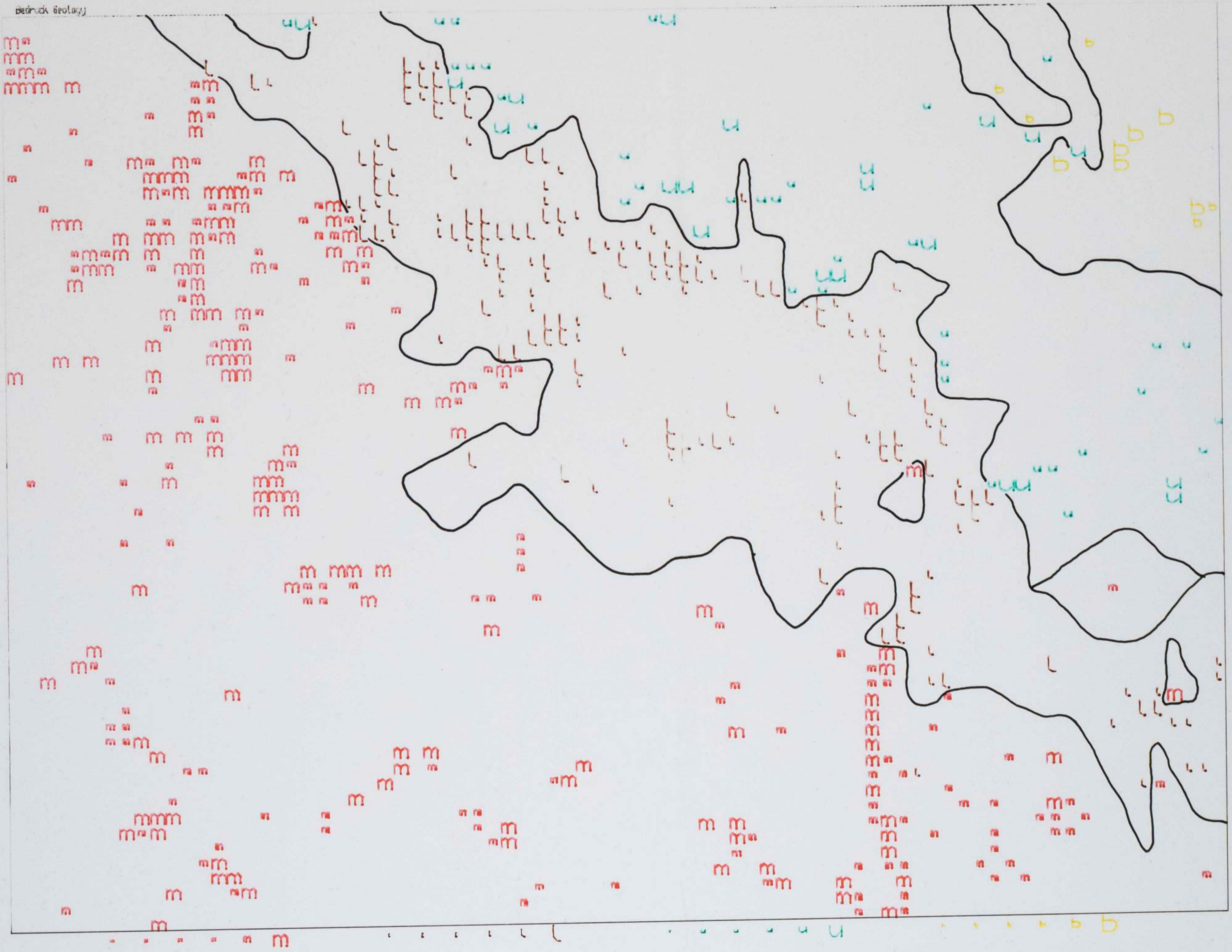


O
F
m
C


Glacial: Ice & Sea

Superficial Geology




Bedrock Geology


Location of wells used

a7 APPENDIX 7 of

A Cartiometrically Orientated Database for the Birmingham Area

by J.V.Earthy

A thesis submitted for the degree of doctor of philosophy

University of Aston in Birmingham Feburary 1983

MAP TITLE	TYPE	SYMBOLS / COLOURS / PARAMETERS / VALUES
point distribution	symbol fine red	sizes indicate either: 1 , 2 , 3 , 4 , 5 or ≥6 boreholes per 100m square symbol segment (All size ranges are taken as small-to-large order)
bedrock stratigraphy	symbol	red 'm' Keuper Marl brown 'l' Lower Keuper Sandstone green 'u' Upper Mottled Sandstone yellow 'b' Bunter Pebble Beds The symbol sizes have the same meaning as in the Point Distribution map.
drift stratigraphy	isopleth	brown 'a' Alluvium red 'r' River Terrace blue 's' Sands & Gravels green 'u' Clayey deposits 's' & 'u' are Glacial NB not as I.G.S. classif. size range indicates: data 800 - 400m distant 400 - 200m 200 - 100m one real borehole } per square >1 real borehole } per square
Glacial deposits	isopleth/ isopachyte blue	The symbol indicates the principal component of the sediment as: 'c' coarse 'm' medium 'f' fine 'o' organic size ranges indicate: 0 - 1.25m thickness 1.25- 2.50m 5.0 -10.0m 10.0 -20.0m >20.0m
Recent deposits	isopleth/ isopachyte green	The symbols and sizes are as the Glacial map.
Made Ground deposits	isopleth/ isopachyte red	The symbols indicate the quality of the commonest man made deposits in the segment. The range is from 'g' good (e.g. ballast) through 'f' fair & 'p' poor to 'b' bad (e.g. refuse or chemical residues). Sizes are as the Glacial map.
Level of the present day surface O.D.	contour brown	Taken from the Ordnance Survey 1:10,000 maps of the area. The contour interval is 10 metres The range is from 70 to 140 metres O.D. The number of contours is seven
Level of the lowest bedrock O.D.	contour blue	Taken from the borehole data for the area. This map uses all the available data to produce a smoothed deepest predicted rockhead relief surface. The contour interval is 10 metres The range is from 60 to 140 metres The number of contours is eight
Hazard or summary map	isopleth	The colours indicate the relative difficulty of development of each 100m square area. The range is from blue, which indicates that there are no forseeable problems, through green and yellow to red, which indicates a high predicted site development cost. The reliability of the map is indicated by the symbol size. This is scored in the same ranges as the drift geology map.

BIRMINGHAM BASE MAP FOR THE HAZARD MAP SET. SCALE 1:25,000