Characteristics of a national topographic digital database

Timothy A. Adams

University of Durham

Department of Geography

CHARACTERISTICS OF A NATIONAL TOPOGRAPHIC DIGITAL DATABASE

A dissertation presented in partial fulfilment of the requirements for the degree of Master of Science in Spatial Data Analysis in Geography.

Timothy A. Adams B.Sc.

Graduate Society

September 1979

Abstract

Large scale topographic information has been supplied for selected regions of the country in a digital form by Ordnance Survey since the mid seventies. Coverage in Great Britain at the present time is still rather limited. It is the intention of this M. Sc. thesis to investigate the reasons for a digital approach and hopefully gain an insight into the characteristics, benefits and disadvantages of this alternative communication technique.

The author wishes to express thanks to the Director General of the Ordnance Survey for the supply of software and topographical information, both graphical and digital, for use in this project. Acknowledgements should also be made to the Social Science Research Council for providing the finance for the year's M.Sc. course held at Durham.

List Of Contents

The annexures to this thesis can be referred to in an accompanying line printer formatted file

Chapter One

Introduction

The Ordnance Survey has functioned as the central survey and mapping organisation in the public sector in Great Britain for almost two hundred years and has been continually involved in the production and maintenance of the national topographic database. As the then Secretary of State for the Environment announced in Parliament on 19th February 1973, one of the aims of the Ordnance Survey is to...

> "produce and maintain up to date basic surveys at 1:1250 for major urban areas and at 1:2500 or 1:10000 for the remainder of the country."
>
> (Harley 1975)

By 1978 there were some 53,316 sheets at 1:1250 scale in the urban areas and 164,461 sheets at 1:2500 scale for the majority of the rural areas (Thompson 1978). In addition to the publication of these basic scales data, the information can be utilised in the production and revision of the derived mapping scales - these being 1:10000 (although some constitute the basic scale for remote regions of the country) and 1:25000 (Searle and Waters 1979).

The Distinction Between 1:1250 and 1:2500 Mapping Sheets

Map sheets at 1:1250 scale cover approximately 5.7 per cent of Great Britain, whilst the remainder of the region is mapped at a basic scale of either 1:2500 or 1:10000. The 1:2500 coverage constitutes the basic scale information for almost 76 per cent of the country.

The criteria governing which basic scale to adopt within an area are based on population figures although there are no hard and fast rules so that the system can remain flexible. Harley (1975) points out that for an area to be mapped at 1:1250 scale, the present rule can be flexibly interpreted to include all towns of 20,000 inhabitants or over and which would require a block of forty or more maps for complete coverage.

The format of each 1:1250 plan is 400mm x 400mm which represents a ground area of 500m x 500m. Surface detail within towns is highly complex and great importance is attached to urban map coverage at large scale since these are usually areas of rapid change and high land values. Although the 1:1250 plans involve a certain proportion of feature generalisations and omissions, the information that they do contain is well presented and standardised by a series of rules adopted by the surveyors and draughtsmen.

The 1:2500 scale mapping series provides basic scale coverage for most rural regions within Britain. Although the sheets have a larger format, 400mm x 800mm representing a ground area of 1km x 2km, they do share many of the attributes of the 1:1250 series. However, the reduction in scale does imply that the minimum thresholds above which a feature can be represented should be larger. Harley (1975) points out that there is a basic, although flexible, rule governing this criterion. The rule states that features have to cover an area of 8m x 8m or greater to be shown in plan. In practice however, this lower threshold is often increased for the 1:2500 series. A typical example could involve the representation of small buildings below the threshold limits which are still included because of their specific topographical importance.

The Mapping Programme

The production of a large scale mapping programme is costly in terms of finance, administration and time. A large percentage of OS resources are devoted to large scale work. Some 800 people are employed as surveyors to provide the input to the system and there are circa 320 cartographic draughtsmen based at the headquarters in Southampton involved with the production of the fair drawings.

The adoption of a conventional map production process has provided the national topographic database in a graphic form on a paper based medium. Attached to such a database must be a sophisticated management system and provision for adequate revision processes. The Master Survey Document (MSD) is the transparent, stable base medium which is used in the field for the inclusion or amendment of any topographical changes which may occur.

Two services provided by OS for users requiring up to date information are SUSI (Supply of unpublished survey information) and SIM (Survey information on microfilm). The MSD can be copied at any time to produce a SUSI copy. After fifty house units of change have accumulated on the MSD, it is microfilmed to produce SIM. The resulting microfilm copy can then be despatched to all OS agents to supersede ~~superceed~~ the old version which is destroyed.

After three hundred house units of change have accrued on the MSD, it is returned to the Southampton HQ for a new edition to be produced. The current system has been successful although there are several drawbacks (Thompson 1979, Simmonds 1978, Marles 1971) some of which are specific to the needs of the larger users of topographical information.

Perhaps the largest users of Ordnance Survey data are bodies involved in:

Local and Central Government
Public Utilities
Education and Science.

Thompson (1978) points out that this type of user requires, in the short to middle term, the graphic applications of the data. However, very often these map sheets contain much more information than the users actually require. Indeed this additional information is often disadvantageous in that it encumbers the actual data which they hope to extract. Gardiner-Hill (1972) explained, for example, how the Electricity Boards tended to trace from the originals a series of new maps specific to their requirements, depicting only the features which interested them. On these skeleton base maps they could then overlay the information they required to show unburdened by the additional features which were not required in the first place.

This approach has been adopted by the GPO too. The decomposition of the OS graphics has enabled the production of data stores for the telephone network. Indeed all Authorities responsible for the distribution of Public Utilities require similar skeleton maps although the overlay of information from one Authority with that of another is difficult since they all have slightly different specifications suited to their requirements.

A major disadvantage of this approach has involved the revision of the data stores. Since the production and maintenance of each skeleton map with its accompanying overlays is far from cheap, the operational lifespan for each graphic has tended to be longer than is desirable. Essentially, the revision of such information would have been necessary when there was either a significant change in the Authority's network or when OS themselves published a new edition base map. This has not been the case in practice. On average, the GPO produce new edition skeleton maps every ten years.

This undesirable situation stems largely from the inadequate format of the topographical information supplied by OS to these users with specific requirements. It has become abundantly clear that the Central Mapping Agency must provide the data in a form that their customers require them and perhaps the traditional litho-printed hard copies on paper do not completely suit the needs of all users.

The start of the change in policy by OS can be reflected by the more recent increase in the demand for maps supplied as transparencies or in microfilm form. To quote the Secretary of State for the Environment in February 1973:

> "The Ordnance Survey no longer has any general obligation to supply maps at 1:1250 and 1:2500 scales printed on paper. Instead it has full discretion as to the form in which the survey at these scales will be made available. Large scale survey information will continue to be supplied but the Department will take account of the variety of possible techniques of publication and will be free to decide how best to produce and market it having regard to the needs of users, income and proper economy".
>
> (Harley 1975)

The Introduction Of Digital Techniques

Traditionally topographical information has been thought of in its cartographic form. Essentially any map sheet can be looked upon as a series of points - some remaining independent and others being linked together to form linear or areal features. Each point, therefore, can be represented in a cartesian coordinate system having four attributes - x, y, z and w. The x, y and z coordinates fix the position of the point in three dimensional geographical space whilst the fourth, w, enables an interpretation through some explicit convention of what the point represents. This w coordinate could represent a bench mark or the corner of a building for example. In taking such an atomic view it is possible to convert a map sheet from its graphic form into a digital equivalent whereby it can be stored and processed by a digital computer.

The supply of topographical information in computer compatible form to users who have access to the appropriate computing facilities has great potential. With the suitable generated software the benefits of a digital format far out weigh the traditional approach:

- selected feature types can be retrieved and plotted thereby eliminating the need to produce labourious skeleton maps.

- specific map areas suited to the requirements for each user can be retrieved and plotted thereby eliminating the need to use large plans for a very small area under investigation.

- extending the previous ideas further, adjacent sheets or parts thereof can be concatenated ~~together~~ thereby providing a solution to the traditional problem of areas lying at the edges of two or more map sheets.

- within reasons, the information can be drawn at any scale. This facility may be advantageous for certain purposes although the cartography may not be too elegant. This could be improved if a scale dependent feature selection routine is included in the software.

From a user's point of view, the advantages of map data in their digital form far out weigh those of the conventional hard copy counterparts. He has the means of drawing the map suitable to the requirements for some particular purpose and is not restricted to the acceptance of what the draughtsman has designed as a compromise between the many conflicting requirements.

The ease of data revision within a digital format is perhaps one of the system's most fundamental assets, not only from the Mapping Agency's point of view but also the customers'. Features can be added, deleted or amended very easily without the need to create the entire sheet again. Since the user's data (for example the Electricity Board) are kept independent of the base map data and are overlayed when required, there is little need for major structural changes within the base data at each revision stage.

The Feasibility Of A Digital Mapping Programme

The Ordnance Survey professional paper number 23 by Col R. C. Gardiner-Hill (1972) provides a good outline of the initial studies which OS attempted in the early seventies to discover the feasibility of designing and maintaining a digital topographic database. They were able to show with their fairly low order graphics peripherals the practicability of handling the basic scales data in a digital format and subsequent regeneration of hardcopy maps to the conventional standards of accuracy.

In 1973 a Ferranti Master plotter, with its light spot projector drawing head, was obtained providing the precision hardware required to produce the final drawings for the production environment. Harley (1975) states that after just a few months, by the end of 1973, a total of seventy nine sheets were available on magnetic tape, (46 at 1:1250, 33 at 1:2500).

According to present figures (Thompson 1978), there are a total of 217,777 maps in the basic scales series, of which 38,099 (some 17.5 per cent of the total) fall in the digital areas. These areas have been selected largely due to the interest expressed by potential users. At the time of Col Thompson's paper in late 1978, some 6,643 of these sheets had been digitised which amounted to 17.5 per cent of the total in the digital areas or 3 per cent of the total number in the series. Thompson estimates that, with present workrates, it will take circa twenty years to complete the 1:1250 scales and a further thirty years to accomplish complete digital coverage within the 1:2500 series.

The Costs Of Digital Map Production

Many in house studies have been undertaken at OS to determine the nature of the costs incurred in a digital map production flowline. Col Thompson, in his 1978 paper, gives much attention to these cost studies. At the present time, the adoption of a digital environment is far from cheap. As Thompson (1978) points out, for 1:1250 map production, costs are 1.6 times greater than those of the conventional flowline. In addition, the resources required in terms of time and manpower are also greater. A similar situation exists for the 1:2500 digital map production flowline.

The reasons for these increases in resources lie in a number of factors:

although a digital approach attempts to automate the system, at present there still remains a large number of manual processes and human intervention requirements;

the current system involves a large number of processes involving many Production Divisions within OS and delays are often incurred due to poor communications within these Divisions;

any digital mapping scheme requires additional checking to those in its conventional counterpart. The correctness of data content and accuracy is of paramount importance and the lack of interactive facilities within the present system suggests a slower edit stage than is desirable.

Although the digital service currently offered by OS is still at a relatively early stage in terms of coverage availability. It should be possible to begin to generate at least some cautious estimates on the characteristics and size of the databank on its completion. It is the aim of this thesis to attempt this.

Chapter Two

The Conversion Of Graphic Data To A Digital Format

There are many techniques currently available to convert a graphic image into its digital equivalent to enable computer storage and manipulation. The most simple and straight forward approach could involve the measurement and coding of National Grid coordinates for subsequent keypunching on to cards. For a more advanced approach there are specific hardware devices to achieve these ends virtually automatically. Typical examples include the raster type of digitiser such as those manufactured by Scitex Corp. Ltd. which scans the image in a series of swathes. This approach has drawbacks in that it requires a great deal of post processing to reconstitute the points and lines from the 'spaghetti' which the digitising hardware produces.

An alternative design for the 'intelligent' digitisers are the automatic line followers of which the Laserscan Laboratories Fastrak is an example. This device has the ability to lock on to a line specified by the operator and then stream encode coordinates automatically until it meets some obstruction - possibly a break or junction in the line. The system can then pause and await for operator intervention.

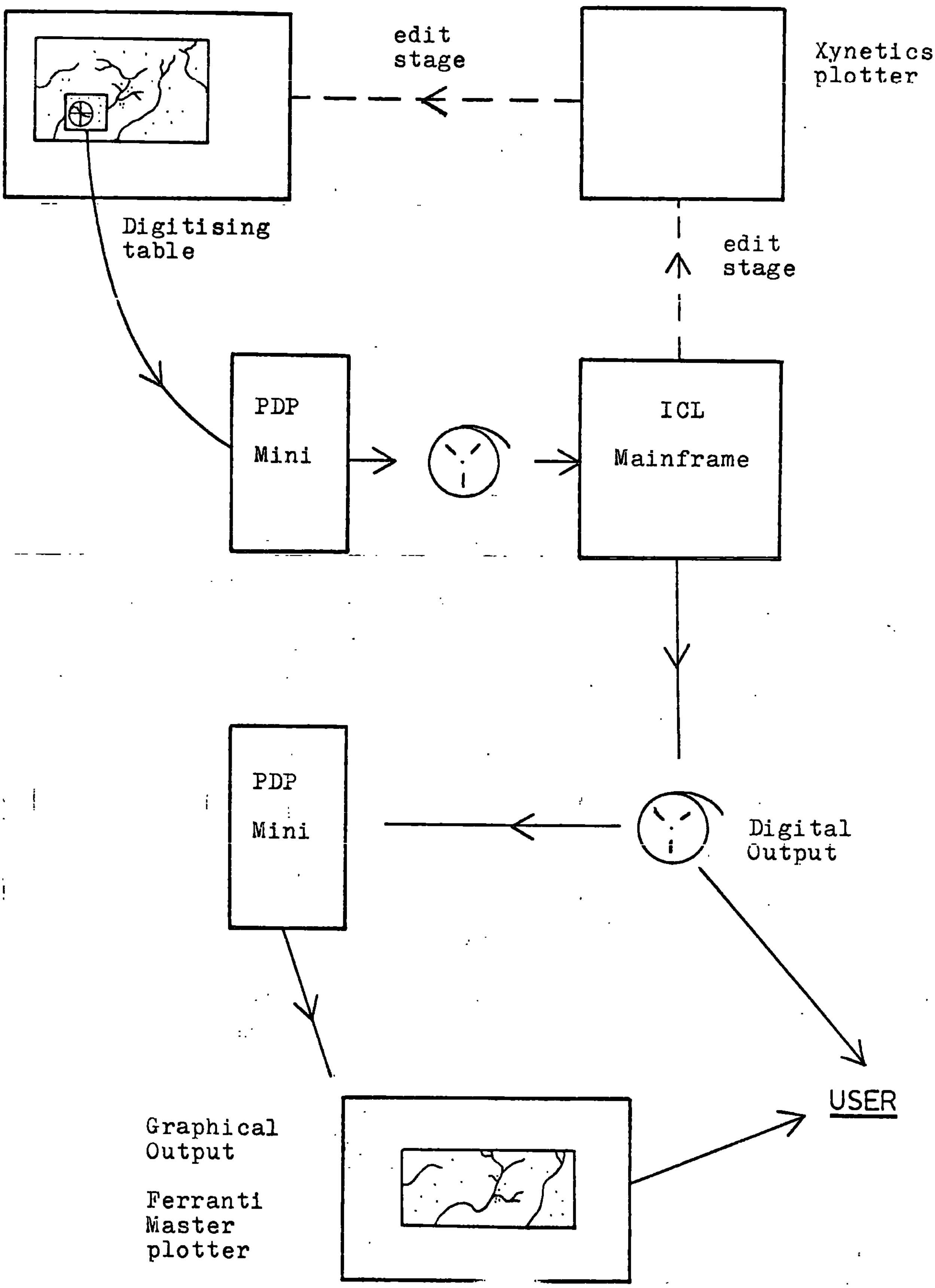
It follows that the more advanced approaches may well be poorly suited to large scale data but could have great potential and provide a much improved throughput with contour type data. At the conference on "Databanks and Digital Mapping" held by the RICS in November 1978, Mr C. Howmans of the Mapping and Charting Establishment (MCE) at Feltham expressed his opinions on the advantages of the system at MCE which incorporates a Fastrak. The hardware requires only one skilled operator (thus saving on eight others) and the throughput is much improved over conventional techniques. A ten to fifteen times increase in contour work and four to five times increase in land use work is not atypical.

Thompson (1978) points out that these sophisticated digitising techniques have been undergoing evaluation at OS although so far, since its inception in 1973, the digital programme has involved the use of the more conventional manual digitising tables. These are characterised by an operator moving a cursor, attached to the table by cable, along the lines to be recorded. The Department has twenty four of the Freescan tables manufactured by Ferranti. Point features are digitised by first placing the cursor in the appropriate menu square and then on the actual feature location. Line features can be digitised in a similar manner although the draughtsmen must use their judgement to decide how many points are required to fix each particular segment. Lines which are straight, as in much of the urban data, present few problems although curves require more attention. In such situations, an offline computer can generate additional points should the operator have underestimated the number required to accurately fix some complex topography.

Figure 2.1 below shows the basic system design operated currently at OS.

Having acquired the necessary set of coordinates required to represent a map sheet, then these data must then pass through another processing stage to arrange them into a suitable order by the addition of special sequence codes. These codes are necessary to enable the plotting program to interpret each file sequentially and act accordingly. It is also at this stage when the textual information for each sheet is appended to each data file. The digitising procedures adopted attempt to reduce to a minimum the requirement of the draughtsman to keypunch information. This should be the responsibility of the normal data preparation staff who are much less prone to making mistakes. By designing the workflow in such a

Figure 2.1 - The DMB System To Convert A Graphic Image To Its Digital Equivalent



way, it is possible to reduce the amount of time that each sheet actually requires on the digitising table.

The Format Of The Digital Data

Most literature concerning automated information systems and digital mapping seems reluctant to differentiate between the terms "databank" and "database". However, the author prefers to adopt a subtle distinction between the two expressions. Essentially the term "databank" is taken to refer to a store of completely unrelated data requiring a sequential search to discover any interrelationships whereas a "database" contains interrelated information which is cross referenced by some master index and a series of sub-files.

The current digital mapping programme operated by OS must be considered to be a Land Information Databank because of its unrelated data characteristics. These data are supplied to the customers in a format suitable for the current OS software to interpret.

This customer format, adopted for data interchange on all magnetic tapes, is known within OS as DMC format. Every client purchasing digital data is issued with the OS documentation "Ordnance Survey Digital Data On Magnetic Tape" to explain the characteristics of this format.

A FORTRAN readable character format has been adopted to ease the problems of data interchange between computer installations. The computer compatible tapes can be supplied in a wide variety of options to suit the needs of all customers. Both seven and nine track tapes are available with optional packing densities and character codes. These options are all outlined in the OS documentation.

Two types of code are involved in the make-up of

DMC format. Primarily there are a special series of sequence codes identifiable as negative integers which enable the users to interpret any subsequent information more easily. Secondly there are the conventional feature codes - the w constituent of the features as described in Chapter One. Appendices A and B show the breakdown of these two types of code and how they can be used in conjunction to form a cartographic databank. Appendix C shows a sample of DMC format for a hypothetical example containing one 1:1250 map of four features; these were extracted from "Ordnance Survey Digital Data On Magnetic Tape" supplied by OS.

The Regeneration Of Graphics From The Digital Information

The Systems Branch of the Department has developed a number of software packages to interpret the data in their DMC format. As far as the customer is concerned at the present time, the plotting program D09 is perhaps the most important.

The software has been written in standard NCC FORTRAN and is therefore virtually machine independent. Its use and options are outlined in the "D09 Program Description" supplied by Ordnance Survey. In order to operate in the manner in which it was originally designed, it requires four peripheral channels:

Channel 1	-	Input Plotting Parameters
Channel 2	-	Input Digital Data in DMC format
Channel 3	-	Output to line printer of the feature listing (optional)
Channel d	-	Output of binary instructions to drive the plotting hardware. (The channel number is installation dependent).

The plotting parameters attached to channel 1 are an additional series of codes that are kept completely separate from the topographical information. Appendix D tabulates these parameters as outlined in the DO9 program description supplied by OS.

It is these codes which allow the user to request specific options leading to the type of plot that he requires. These parameters can be read from card or paper tape medium or, as implemented at Durham, from a disk data file.

The careful selection of the appropriate parameters enables the types of manipulation which the National Mapping Agency initially intended for their digital data. A typical requirement could be the plotting of specific areas of a map sheet, the whole sheet or indeed the concatenation of two or more sheets. ~~together~~. It is codes 01 and 02 which specifies the details of such a requirement to the plotting software. Code 04 allows the user to assign a specific plotting scale. Thus although a digital sheet itself may quote the default plotting scale intended for that sheet, the user can override this and provide his own requirements.

Additionally, it is these parameters which allow the type of features which are to be included or rejected within the plot thereby enabling the development of the skeleton type basic scale map required by so many users.

The rejection of specific features within the plots is possible with any of the following approaches:

(a) By Feature Code

this allows the suppression of all of a certain type of feature - for example, all textual information which has OS feature code no. 28

(b) By Name Category

this allows the suppression of text specific to some particular classification - for example, all text which is classified as a sheet name or perhaps all archaeological names.

(c) By Serial Number

all features in a map sheet are serialised by a specific number of ascending order from start to finish. Provided the user is aware of these, then features can be suppressed from a plot by their unique keys. Control parameter 17 allows these serial numbers to be included on a plot if desired.

(d) Selection By Serial Number

this is the alternative to suppression by specific serial number. It allows the specific selection of features to be included within a plot by their unique keys.

Other control parameters as outlined in Appendix D allow the rotation of plots, inclusion/suppression of any marginal data, titles, etc. Since the topographical data are separate from the control parameters, they can remain unchanged. It is the job of the manipulating software to extract the type of information which the user requires from the databank which is unaltered. These ideas can be carried further when a user wishes to overlay his own information. Since these data can be stored in a separate file, then these too can remain unaltered.

The emphasis thus far has been on the manipulation of the resulting graphic. The current OS production flowline should be looked upon as merely a National Cartographic Databank. This is a splendid achievement on the part of the National Mapping Agency and has made possible a great many more applications for the land information which would not have been possible in its graphic format. However, such a system has much more potential than merely the supply of cartographic information. Since the magnetic tapes supplied by Ordnance Survey contain environmental information then why should these not be the resources for many other applications ? Already cadastral information is held by H. M. Land Registry on the basis of large scale plans and a digital environment at the Land Registry could improve efficiency immensely.

The Present Limitations Of The Production Tapes

The current production environment at Ordnance Survey supplies data for topographical information although it is unstructured for the many applications for which it could be utilised. Why the wrong format ? The production digitising procedures made good sense for the preliminary objectives for the digital mapping programme at the time of the Gardiner-Hill paper (1972). Since remapping from this new data source was the only objective, then a number of points can be made regarding the basic characteristics of the digital data.

Since the storage media adopted for this information are magnetic tapes, which are a form of sequential access memory, each map sheet can be looked upon as a sequential access file in data processing terms. As explained earlier, these maps or data files comprise a series of points which can be plotted in any order. Of course those points suitably linked to produce line or areal representations do require a fundamental continuity within each feature, but the order in which each feature is plotted will be irrelevant. Provided the file is entirely processed, the resulting graphic will appear exactly the same for any sequence which the features may take within the file. In addition, within the terms of reference set out by Gardiner-Hill, there is no requirement for lines within the data structure to be cross-referenced. Two lines which constitute the opposite sides of a road's casing, for example, could be drawn at very different times but this is irrelevant for plot reproduction.

The format of the production digital sheets therefore is inadequate for the databank to be considered as a source of land information from a cadastral point of view. Mention should, however, be made of a number of advantages which this digitising procedure creates. The benefits of such a format can be cited with an example such as a block of terraced houses - a common occurrence on the 1:1250 plans.

Although a Cadastral Information System would look upon these as a series of individual land parcels, to the computer plotters, as a direct result of the digitising strategy adopted at Southampton, they appear as perhaps five points which when joined together at some stage within the file produce the outside walls of the block. It could well be that it is at the other end of the file when a series of paired points generate the boundaries between each dwelling. Figure 2.2 shows this example graphically.

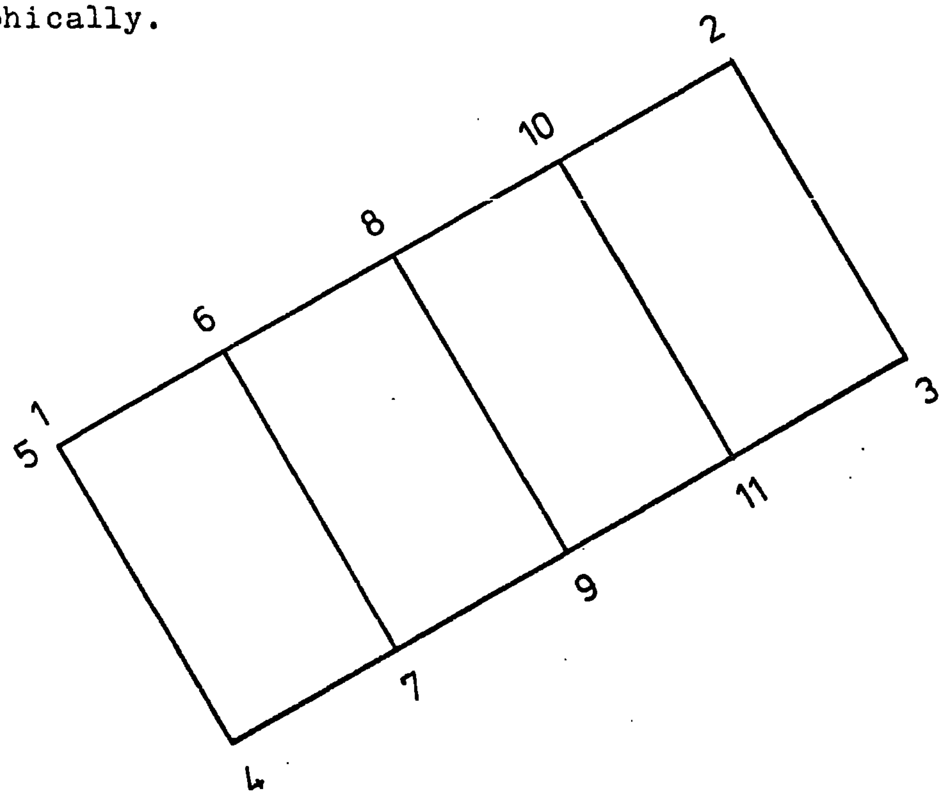


Figure 2.2

The advantages of this approach are two fold: it follows that no line needs to be stored, and hence plotted, more than once as would be the case should each individual parcel be digitised and serialised independantly. Should this be the case, in this example each parcel would require five x and y cartesian coordinates in addition to the three times increase in the necessary sequence codes which DMC requires. Secondly, there is no problem in ensuring that the house frontages are straight since it must obviously follow that only two digitised points (1, 2 in the above example) can only generate a straight line.

The paired points representing the boundaries could well present other problems. These points need not necessarily lie exactly on the frontage, this will depend on both the resolution of the Freescan digitisers, the conscientiousness of the operators and the accuracy of the plotters. Needless to say this possibility is never evident on plots at the designed plotting scale but could appear on graphics at very much enlarged scales. Annex 8 gives such an example. A fifty metre square has been enlarged from digital sheet TQ 6794NE (square 3,6) and has been plotted at a scale of 1:75. It can be seen in this example where there are many occurrences of lines which are not collinear when they should be. Of course, it should be recalled that this area has been enlarged over sixteen times which is undesirable when the data are designed to be accurate only to 1:1250 standards.

Annex 9 shows a much more desirable property of the DMC production format. Sheets NZ 2551SE, NZ 2551NE, NZ 2551SW and NZ 2551NW have been concatenated ~~together~~ and plotted at a reduced scale of 1:2500.

Within the given terms of reference outlined so far, the current production flowline operates well. Should more complex manipulation of the information be required, for more analytical tasks, then the present format of the magnetic tapes provide severe limitations. The cadastral problem already cited is the classic example of the inadequacy of the data in their unstructured form. This occurs directly as a result of the independence of each of the line segments within the file.

The Need For A National Topgraphic Digital Database

The potential that a topographic, as opposed to cartographic, digital database could offer is widespread. The role it could play in automating H.M. Land Registry is perhaps the major advantage that a Land Information

System could provide.

On the whole, any areas of study which require an interface with the information on large scale plans could have scope for becoming partially automated or at least more efficient with digital processing techniques. Local Authority property records could be efficiently handled on line. Land Information is a major resource of any L. A. and is required by all levels of officer in order to carry out their day to day operational functions. This concept has been kept in mind by the development team of the LAMIS system at Leeds (Harrison 1978). Should the regeneration of individual land parcels be possible from the OS data, then the UPRN (Unique Property Reference Number) concept of the Local Authority Management Information System could be much more flexible.

As an alternative example, the CLUMIS system developed by Ferranti-Cetec Graphics Ltd., in Edinburgh can be used in collaboration with the OS data for:

(a) derived mapping at smaller scales (after generalisation)

(b) generalised outlines of topographical features if selected OS data are used and the CLUMIS interactive facilities envoked.

On the whole, much benefit will be gained by all these systems if the topographical information is coded in cadastral or spatial units rather than in topographic terms.

A Land Information System operated by Ordnance Survey need not necessarily be restricted to cadastral uses. There is no reason why the topographical database should not be coordinated with: Land Use studies, Pollution Monitoring, Transport Studies, Terrestrial Ecology as well as those data provided by the Office of Population Censuses and Surveys.

The Need For Environmental Information Management

"Who should provide the management for a Land Information System ?". This question has generated much controversy in recent years. Rhind and Trewman (1975) suggest that reasons for these arguments could possibly stem from the lack of definition over the role which a cartographer should play in the mapping industry. It may be argued that a cartographer's job is to produce maps, he should not involve himself with the contents and characteristics of a topographical databank. However, the authors pose the question "has it not long been the role of the cartographer to involve himself with certain aspects of data acquisition and organisation ?". The answer to this question could certainly produce objections from those people who hold the previous narrow view of the cartographer's role.

Perhaps a more useful implication is suggested by Dale (1977) who looks upon the expertise of a land surveyor as best fulfilling the qualifications necessary for a land information manager. Dr Dale points out that along with measurement, mapping is a principal component of land survey. Consequently a land surveyor should be held responsible for the total process of acquiring, processing and presenting topographical information.

A solution to this apparent controversy could be derived if the two areas of land surveying and cartography were considered as becoming one multi-concept discipline. The implications of automation in terms of data acquisition, manipulation and presentation could greatly benefit from the expertise and experience of geographers, surveyors and cartographers alike. It is of paramount importance that the management for an environmental information system should be drawn from all three types and of course the resulting team must then make it their duty to become familiar with computer systems and applications.

As such, therefore, the Ordnance Survey is well suited to operate the database. The department has served as the National Mapping Agency for two hundred years. Over this time it has obviously acquired great skills and expertise in the handling of topographical information. The change from a graphic to digital format alters their role only in the necessity of understanding completely the vast potential that this change offers. It remains the responsibility of the users to inform the managers of the types of service which they hope to obtain from these new systems.

The Importance Of Knowing The Characteristics Of The Database

Having outlined the basic structure of the database in its present format and the various uses to which it can be put, it now becomes necessary to consider the characteristics of the data and their effects on the users. An a priori understanding of the feature codes adopted by OS is of fundamental importance to any user wishing to take full advantage of the data in their digital form.

A complete digital map sheet is built up of many differing features with their appropriate numerical code. If a user is to plot the exact information he requires then he must be fully aware of the coding procedures adopted at the digitising table.

In order to gain an insight into these procedures, a series of experiments were carried out to break down a 1:1250 plan into its component parts. These investigations were carried out by the author with Durham sheet NZ 2642NE. It is known from subsequent statistical analysis that plans of this nature have over two thousand features in all with an excess of thirty classified types. Software was written for the investigation to break down the sheet into its atomic parts, the source code of which is listed in Annex 1. This software was designed to read a file in Ordnance Survey DMC format and generate a series of subfiles for each of the feature types. The DMC format is preserved throughout in order that the D09 plotting software can subsequently process each subfile in turn and generate graphic output from the breakdown. The plotting scale has been reduced from the default 1:1250 to 1:2500 to keep the results compact for display purposes. These results can be seen in Annex 7.

The results of the investigation show quite an unexpected series of graphics. There appears to be little coordination of lines within each plot. For example, all house frontages which are coded 3 (Building - other), appear quite separate from their boundaries which are coded 30 (Fence, Wall etc - non road). This example provides good evidence of the unstructured characteristics of the DMC format. Additionally it explains quite convincingly why the user must understand fully the coding procedures adopted. If, for example, he requires a plot of all building information and suppression of everything else, then it may not simply involve selecting all features coded by the classification 'building'. This was attempted also during this exercise; the results appear at the end of Annex 7. It can be seen that equally poor structured plots develop since the graphics are missing certain other features which the user may not be aware that he must select. These limitations could also severly affect the production of skeleton maps for other types of information. Annex 7 shows an example of a skeleton map for road information although this too is poorly structured.

This road based skeleton map shows a second severe limitation of the current data content. The Durham sheet under scrutiny has a railway viaduct in the south eastern corner and since this information has been suppressed, all the road information which passes under the viaduct, including the centre line, is broken. These two examples help to show possible drawbacks in the current data format from a cartographic point of view in addition to those already discussed from a cadastral point of view.

The investigation has attempted to show that if a user requires a specific skeleton map, he must know precisely which features to extract and which to suppress. This information cannot readily be obtained from a mere glance through the feature codes list, there is a need for an à priori understanding of the coding strategy.

Two other characteristics should be pointed out which are not immediately evident from any literature so far available on the subject. An example of gross enlargement of plotting scale has already been discussed with an example given in Annex 8. Perhaps the resulting poor closure of features should be expected at this over enlarged scale. However, these misclosures may have serious consequences for more analytical tasks. A typical example could involve attempts to compute the area of features, or volumes if the contours of a 1:10000 map are included. Certainly the user must be aware that he requires to specify a series of threshold tolerances in his software in order to handle these misclosures.

Misclosures of this nature are to be expected when the digitising process is not followed by post processing to ensure point collinearity, a technique ignored by OS.

A second characteristic worth noting can develop when two or more sheets are concatenated ~~together~~ for plotting purposes. As one would wish, there seem to be no problems in ensuring the continuity of features at sheet edges although as the example in Annex 9 shows, there is the disadvantage of repeated textual information. This can easily be resolved by suitable suppression of the repeated text but perhaps users are not immediately aware of this characteristic and certainly it will be necessary to produce an initial test plot. No doubt a cathode ray display unit with graphic facilities will benefit the system here.

Attempts have been made in this chapter to introduce the development of the digital database and consider the potential which such an information system might have. Possible limitations have been shown which may be resolved with the development of suitable software. Finally it has been suggested that perhaps potential users are not fully aware of the possibilities which such a system could offer and certainly any literature currently available on the characteristics of the data seems to be scarce.

Chapter Three

The Aims Of The Study

Maps at the 1:1250 scale cover approximately 5.7 per cent of Great Britain (Harley 1975) and provide the basic scales data for the urban regions of the country. Each sheet represents a ground area of 500 metres square. The remainder of the country is mapped at 1:2500 scale although in remote mountain and moorland districts, the largest scale reduces to 1:10000. These three published scales provide the basic national map series. The conversion to digital storage for these data is no mean task as pointed out by Thompson (1978).

After some five years of digital map production at Ordnance Survey, some 3 per cent of the total basic scale maps have been digitised. The present policy, however, is to convert map sheets which lie only in specific areas. These proposed digital areas represent 17.5 per cent of the total in the series and in late 1978 some 17.5 per cent of the sheets in these areas had been digitised (Thompson 1978). The proposed digital mapping regions were designated as a direct result of the response and interest expressed by potential users in automated mapping applications.

The amount of information stored within the database on its completion is expected to be fairly extensive. The characteristics of these dimensions will be very important from many points of view. Since its inception in the early 1970's, much investigation has been undertaken to discover the uses, management and efficiency of such a system. An area tending to be neglected in the recent past has been a study of the database dimensions both now and in the long term. Even at Southampton, there has been very little in house study to provide any estimates.

One of the chief objectives of this thesis has been to investigate the possibility of general statistical analyses of the data with the view to producing some projected estimates of the dimensional quantities for the completed system. These estimates will be beneficial to the mapping industry and land information managers alike.

The decision to form this collaborative project with OS was taken in February 1979. The university environment at Durham is well suited to such a task since the data processing facilities are very good. Thus research can be undertaken with these facilities leaving the OS hardware free to specialise in the development of their production environment.

Durham University is the joint owner of the Northumbrian Universities Multiple Access Computer system (NUMAC) which is based on two IBM central processors - a System 360/67 and a System 370/168 with virtual memory. The latter machine operates under the MTS (Michigan Terminal System) executive which is a time sharing multiple access system allowing one hundred or more terminals to be on line at any one time. This machine was to be used for the software development and data manipulation requirements for this thesis.

Data interchange between the OS and NUMAC installations presented few problems owing to the wide variety of tape formats offered by the Mapping Agency. Nine track computer compatible tapes were adopted, with a packing density of 1600 b.p.i. storing the DMC format as EBCDIC characters.

The sheer magnitude of 4502 digital sheets presently available (Thompson 1978) prevented the possibility of OS supplying the current database to Durham. A carefully designed sample of the sheets available was required, therefore, to provide the best series of estimators that such a study could offer.

The Design Of The Sampling Procedure

The design and implementation of a spatial sample has long been of major concern for Quantitative Geographers. In general, they take a sample if they wish to generalise a population or more usually because they cannot handle the whole population in terms of computing time and data storage. The latter of these two motives is the one which concerns this project.

In the past, researchers have attempted to base their samples on a rigid statistical or probability approach. Berry and Marble (1968) suggest variations on the basic random through to systematic themes. Haggett, Cliff and Frey (1977) point out,

> "there can be, essentially two approaches to sample design:- purposive or 'hunch' sampling where individuals are selected which are thought to be typical of the population as a whole, and probability based sampling where the individuals are drawn on the basis of rigorous mathematical theory."

With the latter of the two approaches, once the design is adopted, individuals are drawn according to the established set of rules.

Ideally, this investigation would have greatly favoured a probability based design since knowing the fundamental set of rules governing the selection of the sheets would have enabled confidence limits and statistical inferences to follow far more convincingly. However, a 'purposive' sampling design had to be adopted for a number of reasons:

(a) digital coverage of the country is fairly sparse at the present time;

(b) the designation of the areas which are to have digital coverage has followed no formal rules - they have developed as a result of user interest;

(c) for administrative reasons, the OS could far more easily handle the supply of blocks of information as opposed to random sheet selection given that the number required was of the order of 500 sheets.

Since a 'purposive' approach very much depends on any ideas held by the researcher in the outset, great care is necessary to ensure that the selection of the individuals does not produce a biassed summary of the population. The digital areas are fairly well scattered throughout Great Britain and hence the sample was drawn from as many of these areas as possible. In this way, it was hoped that the results would not be too influenced by local anomalies as may have been the case had the entire data been drawn as a block from few digital areas.

It is probably to be expected that a digital sheet from a dense urban area will have different characteristics from either a rural area or a sparsely populated urban region. In an attempt to investigate the truth of this hypothesis the sampling procedure adopted was based on the ideas of stratification. A sample was drawn from each of the strata at both 1:1250 and 1:2500 scale wherever possible. An additional stratification criterion was included based on the nature of the terrain which the maps provided coverage for. Within the digital areas, there is a great variation in the ground terrain ranging from the relatively flat areas of Birmingham to the much more hilly regions of Derwentside. These topographic differences will probably have some effect on their populations but perhaps more important in this instance, could provide different digital characteristics for the maps. Height information itself is irrelevant on the basic scale 1:1250 and 1:2500 sheets, but it may affect characteristics such as the segment distances of digitised line features for example.

The sampling design adopted for this study, therefore, was not probability based although great care was taken in the design of its structure and the stratification approach ensured that both terrain and population size had valid representations, given the limitations outlined earlier.

The actual sheets obtained numbered 478 in all (210 x 1:2500 scale, 268 x 1:1250 scale). The stratification breakdown is outlined in Appendix E.

The Approach Adopted For The Investigation

Any map sheet in a digital format enables a great many more manipulation possibilities than does its graphic equivalent. Once an understanding of the DMC customer format of the OS digital tapes has been achieved, it is not too difficult to develop appropriate software to carry out a great variety of processing skills.

A study of the characteristics of the topographic digital database as operated by OS can be subdivided into several broad subdivisions. Primarily it is possible to consider each map sheet as an individual, and proceed to generate a series of statistical quantities for each sheet as a single unit. Comparisons can then be made between the absolute quantities which are derived for each sheet.

It is a fairly trivial task to extend these basic concepts, treating the whole of the sample as a single unit. Thus instead of generating a range of statistical quantities for each individual, they are concatenated ~~together~~ to derive a single set of statistics for the sample as a whole.

Finally the individual sheets can be grouped together into their stratification units as outlined in the sample design procedure. Hence a similar series of statistical quantities can be developed for each stratum. This will allow comparisons between each set of results and possibly show the effect of both population and ground terrain on the characteristics of the digital sheets. Of course for comparison purposes by strata, it becomes necessary to weight each set of results as a function of the number of sheets in the strata. This can be most practically achieved by dividing each result by the number of sheets in the stratum. The statistical quantities would then revert from absolute numbers to ratios.

Since these three basic approaches can be carried out for data from each mapping scale (1250 and 2500), then comparisons can additionally be made between the two basic scales.

Types Of Statistical Analysis

As far as the author is aware, there has been little research undertaken involving the statistical analysis of OS digital data. It would follow, therefore, that any tests, no matter how trivial, would be an improvement on the status quo. Rhind (1974) points out the idiocy of carrying out more sophisticated analyses on a data set without first discovering the most simple of univariate statistics for the distribution. The simple statistics developed in this instance involve: absolute counts of features, the distance they represent at both ground and plotting scale, the total number of line segments, the total number of data points, etc. These sorts of data will be hyper-useful for planning, plotting or further analytic purposes.

The ability to achieve statistical quantities of this nature is a direct reflection on the properties which a digital format can offer. The calculation of simple univariate statistics can give the user a familiarity with the map sheets from a quantitative point of view as opposed to the mere qualitative impressions which the visual study of a graphic provides. These basic quantities will give, perhaps, the first indication of the types of feature which a topographic information system will contain. They will show the most prominent features and possibly provide an idea of their relationship with population and ground terrain. Projects of this nature have always been possible with the conventional paper based maps although the manual procedures involved tends to be labourious and hence discourages any interested parties from tackling the problem. Certainly one could not hope to achieve the accuracy that a digital approach can provide given that the accuracy of the data in the outset is to a suitable standard.

From the results obtained for the sample, it is possible to project the values to their equivalent for the entire country. The stratification of the sample should improve the reliability of the results. Suppose we had a sample of three sheets out of 150 in a strata, then this constitutes a 1 in 50 (2 per cent) sample. To project the sample to the one hundred per cent population would entail an enlargement of the results by a factor of fifty times.

The approach is not rigorous and certainly leaves much to be desired but at least it can improve upon the absence of any previous results of this nature. The limitations of the technique are most dependent on a number of basic assumptions:

- Although the arbitrary digital areas are further classified into strata, there is still the underlying assumption of homogeneity of the map sheets within each stratum. This is almost certainly an invalid supposition, since the chances of homogeneity in population and terrain over all the stratum are minimal. Indeed, there are no hard and fast rules governing the nature of the topography in each of these regions.

- The selection of the map sheets within each strata is not random - the sample is the best that OS can provide given (a) the present sheets available in digital format and (b) the huge administrative requirements on the part of the data control section of OS.

- The contents and characteristics of each individual sheet will probably differ considerably from one to another. The structure of the data depends on operator's digitising procedure. Although digital files pass through many edit stages before they can enter the databank for storage and marketing purposes, there may be instances where features have been omitted. Indeed, Thompson (1978) suggests that many land parcels, for example, are not completely depicted on the large scale sheets. This drawback could present further problems should attempts be made to interface the system with H.M. Land Registry.

- Since the digital mapping regions have been designated as a direct result of user interest, it would be fairly truthful to suppose that these are areas where much topographical change is likely. Thus the digital areas themselves should not be considered typical of the country as a whole.

Given that these underlying assumptions are valid is almost certain to produce a biassed sample, i.e. one which will not represent the country as a whole. However, at this early stage in the digital mapping programme, such a sample is the best that we can hope to achieve. It should serve, at least, as an exploratory tool for further, more valid and complex, analyses as more digital information becomes available.

The 478 digital sheets included in the study form a 1.25 per cent sample of basic scale maps in digital areas, although they constitute a 10.9 per cent sample of the digital maps currently in production.

Software Design And Development

Programs for the statistical analysis of the digital data were written for the NUMAC installation in standard NCC FORTRAN. Since the magnetic tapes supplied by OS held topographic information in DMC customer format, the software was designed to enable manipulation of DMC records.

During the initial stages of development, the types of analyses and the nature of the approach were fairly indecisive and hence the software took a modular design. In this way, it became possible to add more 'modules' as the research progressed and as ideas developed. Consequently, one large program evolved which could perform a multitude of operations on the supply of some specific user request. This proved to be far more beneficial than the adoption of a series of separate programs since the latter alternative would have required either repeated code or a more complex and frequent link and compilation process at the run time. The single large package could be stored in its compiled form ready for immediate load and execution thereby avoiding costly compiler runs each time a program run was required.

The Database Held At Durham

Computer processing of the OS data in their DMC form is costly in terms of time and storage. On the whole, any data manipulation or arithmetic operations requiring CPU time are trivial and the drain on the computer resources stems from the inefficiency in the Input and Output (I/O) operations - the data transfer from backing store such as tape or disk into core for CPU manipulation.

Since the topographic data are held on tapes in a character format, the software has to rely on FORTRAN read and write operations as opposed to the faster and more efficient executive I/O. The FORTRAN I/O is much slower since the system must convert the EBCDIC code into a binary equivalent. The system also validates all transfers by means of checking routines which executive I/O does not. Both the coding conversion and the data validation are beyond the control of any FORTRAN program.

This slower, inefficient data transfer produces great disadvantages with the OS information since each record size is small, restricted to only eight characters. (See appendix C). There require to be, therefore, many I/O operations for each sheet to be processed with a correspondingly higher number of conversions and validations.

Needless to say if the data analysis at Durham was to be most efficient, each tape was required to be processed once and only once, keeping I/O to a minimum. The results could then be stored on disk providing a database for rapid retrieval. The database developed and accessed by the software is subdivided, essentially, into three files:

File 1 - on FORTRAN channel 10 stores accumulated statistics for the entire data as a whole. Thus as more sheets are processed, the counts of features, records, etc. increases since this file treats every sheet as an addition to those which have already been analysed. It follows that at any time, it is possible to obtain the status of the data study. This facility is envoked by the request command 'STATUS' and presents the user with a summary of the contents of file 1.

File 2 - on FORTRAN channel 11 stores the names of the map sheets as they have been processed and included in the data study. As appendices A and C show, the DMC format references each map sheet by the N.G. coordinates of the south west corner. It was considered to be more beneficial if this referencing system be converted back to the traditional system as used by OS on their conventional paper based large scale data. The software handles this conversion and stores the results in file 2.

File 3 - on FORTRAN channel 12 is perhaps the most important file in the database. It is here where the statistics for each individual sheet are stored. Thus the software treats file 2 as an index or hub-file and then from this information it can compute the specific addresses of the corresponding information stored in file 3. This enables rapid retrieval of results for any desired map sheet.

The User Specified Commands For The Data Study Software

The algorithms were designed to be self-checking enabling a novice user to envoke the program in interactive mode and supply requests after suitable prompts from the program. A listing of the FORTRAN source can be referred to in Annex 2.

There are eight specific facilities currently offered by the software and these can be requested by the appropriate program command, which are outlined below:

INITIALISE - this initialises the database, ensuring that sufficient backing store is available, that all counters are set to zero and that all files are emptied. It should not be used after results have been obtained unless the user requires to destroy this information and start again.

QUIT - This enables the user to terminate execution at any time.

HELP - this assists a novice user in how to use the program.

ANALYSE - this envokes the routines necessary to statistically analyse OS digital data which are read in on FORTRAN channel 3. The results are output to channel 2 and also to the disk database. A default of 40 map sheets are processed at any one time or until the DMC end of file code (-3) is detected.

<u>RECOVERY</u> - this invokes the routines necessary to access the database and retrieve the statistical information for any user desired map sheet. The routines supply suitable prompts and error messages. Two types of access are available. For the novice user, the system can search for the absolute addresses in backing store of any desired sheet information on the supply of its name. For the more advanced user who may know these addresses, the system can override the search and allow faster data retrieval.

<u>STATUS</u> - this groups all individual sheets into one whole data set and summarises the information on accumulated counts and moment statistics. There is also an option for a listing of all the sheets so far included in the analysis. Of fundamental importance to a user is the knowledge of how many sheets have currently been included - this command provides this information.

<u>STRATIFY</u> - the introduction of strata into the sampling procedure has already been outlined earlier in this chapter. This facility enables the user to group two or more sheets together and present a similar series of results for each stratum. Since the counts are divided by the number of sheets included in the stratum, then they will no longer be absolute but ratios. These ratio data can then enable comparisons between strata.

DESCRIPTION - this is an option enabling the user to interpret the meaning of any OS feature code. There are options for either selected codes or a summary of the entire set (codes 01 to 271).

Attempts were made to keep the software machine independent although in some circumstances the algorithms make beneficial use of some of the helpful facilities offered by the MTS operating system available on NUMAC.

Owing to the modular design of the software, the ordering of the sheets at process time was irrelevant. This requirement was well suited to the format of the tapes received from Ordnance Survey. However, the 1:1250 and 1:2500 data were kept separate to avoid confusion. This enables the 'STATUS' routine to allow comparisons to be made at the two scales.

Chapter Four

Statistical Analysis Of The Basic Scales Data

Any results generated from a project of this kind should be looked upon as an exploratory tool only. The nature of the approach, together with the limited sample size, can provide only cautious estimates for the characteristics of the database as a whole, on its completion in the long term future.

In order that this work can act as an introduction for any subsequent data analysis of the topographical information, the types of statistical technique were kept fairly trivial. The nature of the results is essentially one of counts for the information encompassed in each digital map sheet. The software developed for the purpose presents its findings on the line printer in three distinct sections. These sections were developed as a direct result of the characteristics of DMC (as outlined in chapter two).

Section one gives information on the nature of the special sequence codes encountered during the analysis. These are the particular codes required by the D09 plot program for routing within the software whilst a digital file is being processed. Thus, the information presented in the first section of the results is one of counts for the seventeen sequence codes (-1 to -17). There can, of course, be many redundancies in this information and the software can check for any internal inconsistency encountered at this stage. For example, since code (-1) represents the start of a sheet and correspondingly, code (-2) represents the end of a sheet, then obviously the two counts for these codes must be equal if the software has performed correctly. A similar situation is present for start and end of features (codes -4 and -5), or the breakdown of features into line type (-7) or text and symbol types (-8) with the total feature count (-4).

Section two provides a more detailed statistical breakdown of each feature type encountered during the sheet analyses. For point symbols and text features (which DMC treats exactly the same), there is little information that can be provided other than their frequency of occurrence and the number of points involved per feature code. The latter two pieces of information can again provide a redundant check since they should be equal.

It is the line features which can benefit most from this detailed statistical breakdown. OS handles a line feature as a series of digitised points which when linked together produce line segments. Studies have been made on these line segments in addition to the basic counts already discussed. Each line feature can be looked upon as a statistical distribution of line segments, each being characterised by some specific length which represents either a ground distance or a reduced scale plot distance. As a supplement to the total line length involved for each feature type, the program provides an indication of the nature of the segment length distribution. The maximum and minimum segment lengths encountered (at ground scale) is reported and the moment tests for normality are provided.

Figure 4.1 below shows how skewness and kurtosis measures can provide information on any deviations from normality.

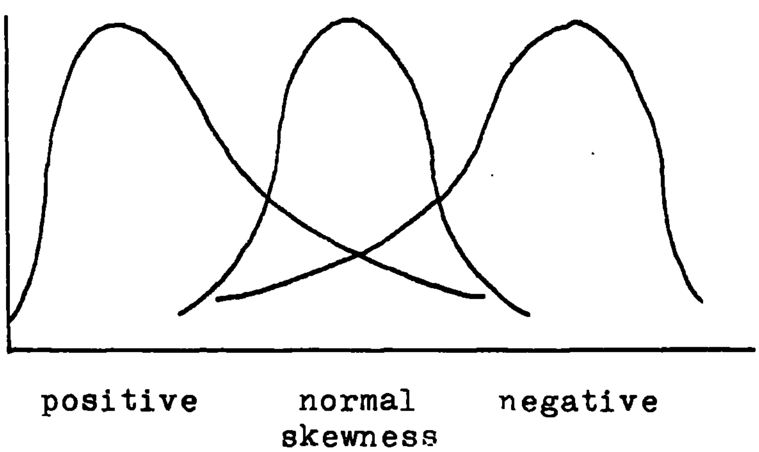


Figure 4.1a
Degrees of skewness in a distribution

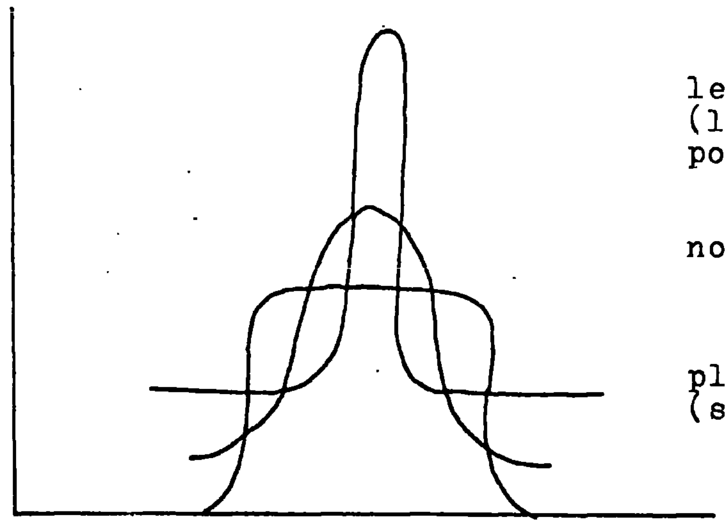


Figure 4.1b - Degrees of kurtosis in a distribution

Having outlined the various distribution shapes possible, it is necessary to adopt some basic measures to summarise them. These measures can be obtained by examining moments about the mean.

(a) The Third Moment - this involves cubing the deviations about the mean to determine any total deviation from symmetry. By summing these cubed deviations, the dimensions become standardised for comparison and it is possible to produce a dimensionless measure by dividing the quantities by the standard deviation raised to the third power (s^3). Cubing the deviations means that the sign is still relevant. Thus a relationship is made between negative and positive values. Should the negatives be dominant, then the distribution will have a negative skew or vice versa.

(b) The Fourth Moment - this involves raising the deviations about the mean to the fourth power, hence sign becomes important - the result is always positive. Since large deviations to the 4th power become much larger, weight is given to the tails whereas the central part remains unaffected. Normal distributions produce a fourth power moment of 3 - the PLATYKURTIC is greater and the LEPTOKURTIC is less. The program avoids this condition by subtracting 3 from all kurtosis measures to zeroise normality.

N.B. Skewness $= \dfrac{\sum (x - \bar{x})^3}{(n - 1) (s)^3}$

Kurtosis $= \dfrac{\sum (x - \bar{x})^4}{(n - 1) (s)^4}$

where: x = segment length
$\bar{x}$ = mean segment length
s = standard deviation of the segment lengths
n = number of segments

Section three of the results provides information on each map sheet as a whole. It supplies counts of the total number of points and line segments and gives an indication of the amount of text included in the sheet by counting individual characters. Of fundamental importance too is the number of records required to digitally store a map sheet in DMC format. Since each record or line of DMC data is an eight character string, in IBM terms this represents a concatenation of 2 x 32 bit words of storage.

Ignored Records And Codes

As the analyses progressed, it became evident that there were occasions on the magnetic tapes where feature codes were detected which could not be categorised by name. Appendix B shows that the feature type classification adopted by OS is coded from 01 to 271 inclusive. However, there were features detected, at scales 1:1250 and 1:2500, which were coded as zero. Since these could not be categorised, it was decided that they should be ignored and excluded from the analyses. The software is designed to report the existence of these ignored features and count how many records are involved in the exclusion. This coding inconsistency arose in the digital mapping system at Southampton when a transfer was made from DMA, the initial test system, to DMB the current production flowline.

Summary Of The Results For The Basic Scales Data

1:1250 Scale

(a) Individual Map Sheet Analysis

For consistency, Durham sheet NZ 2642NE has been chosen as an example of a typical 1:1250 digital plan. This is the sheet which was adopted for the detailed breakdown as discussed in chapter two and outlined in annex 7.

The results of the statistical analysis are shown in part a of Annex 3. The sheet is not of an atypical size for 1:1250 data comprising some 31,115 records, of which 14,789 of these store appropriate x and y coordinates for graphic regeneration. The sheet comprises a total of 12,689 line segments which represents in excess of 49.5 kilometres at ground scale (nearly 40m at plotting scale). By far the largest users of line segments are features coded as number 30 (Fence, Wall etc. - non road) - these accounting for 12.5 km of linework at ground scale. There were 778 features categorised as code 30, involving 3,661 points or 2,883 line segments. The mean ground distance of the segment lengths is 4.4 metres although this conceals vast extremes ranging from 86.38 to zero metres. The occurrence of this minimum distance of zero is very infrequent and arises due to the round-off process undertaken at Southampton when the basic grid squares are converted from binary representations (1024) to their decimal equivalent (1000) for DMC.

The sheet as a whole comprises 31 of the 109 feature types available for large scale usage and possesses 2,100 features in all (1,804 line type and 296 text or point symbols).

This type of statistical information is now available for every sheet held at Durham in the database and can be retrieved very easily. The size of the 1:1250 sheets and feature frequencies does vary, of course, from sheet to sheet although those within urban areas have breakdowns which are similar to the Durham sheet. Feature code 30 forms the bulk of the linework on every sheet although those with many roads do tend to possess more lines than others where routeways are less frequent.

The size of the 1:1250 sheets vary from around 8,000 records to in excess of 35,000. As one would expect, the sheets with most records are located in areas of high urban population density such as Birmingham or Southampton (both in digital areas).

(b) The Entire 1:1250 Scale Sheets Held At Durham

These are the results obtained when the request 'STATUS' is issued to the data study software; a listing of the findings can be referred to in part b of Annex 3. Since there are a total of 268 plans at 1:1250 scale held in digital form at Durham, the figures generated are large but nevertheless the basic trends reflect a similar picture to that outlined by the Durham sheet in part a of the annexure. Feature code 30 has the greatest frequency involving some 505,751 points and 361,016 line segments, thus giving further evidence to the typical nature of the Durham sheet.

At this stage, some 384,197 features had been included in the 268 sheets involving the use of 5¼ million DMC records. It should also be pointed out that 2,245 features had to be ignored since they had been coded as zero. This was a wastage of 73,151 records (or in excess of 585 kilobytes of storage) accounting for 30 feet of magnetic tape (packed 1600 bpi).

On average, the mean distance for most line segments is of the order of 8 metres at ground scale (6mm at 1:1250) although feature code 47 (electricity transmission line) has a mean segment length of 91m at ground scale which reflects the large extents of straight transmission line in the country.

An inconsistency was detected in section one of the 'status' results. The start and end of feature counts do not tally. This was traced to sheet SP 0588SW which has one feature that does not possess an end of feature code (-5). The consequences of such an occurrence are minimal since it is the start of feature code (-4) which has a fundamental role in both the OS plotting software and also this data study software.

(c) Stratification

Part c of annex 3 gives the results of the statistical analyses based on the six strata adopted in the sample design. Since these data are divided by the number of sheets in each stratum for comparison purposes, they are now measurements on a ratio scale. The map sheets located in the dense urban regions tend to be much larger in data content (eg Birmingham averages 25,377 records per sheet) than the lighter dense regions (eg Billericay averages 11,511 records per sheet). These figures obviously reflect the more complex topography since other counts too are correspondingly higher in the high density urban areas. The Birmingham stratum has on average 1,932 features per sheet whereas Billericay has a much lower figure of 773.

The general trends stemming from the detailed feature breakdown remains the same with code 30 accounting for the most sizeable proportion of the linework.

The Southampton region, which is categorised as the dense urban / hilly stratum seems to account for a large number of the zero length minimum line segments discussed earlier. This may possibly reflect on the early development of the digital mapping programme, DMA, involving mapping for the Southampton area.

The ground terrain appears to have little effect on the segment lengths contrary to the expected. For example, code 21 which is fundamentally important so far as routeways are concerned (carriageway road pecks) have the same mean segment length in both the Tamworth and Chester-le-Street strata. Both of these strata have average density population although one is hilly whilst the other is flat. This trend virtually follows for all line features, indeed the Tamworth and Chester-le-Street strata are almost identical in all respects. They would probably differ most at the 1:10000 mapping scale which includes height information in the form of contours. A vast increase in the number of contours is bound to affect the sheet size in digital terms.

The information enclosed in parts a, b and c of Annex 3 is immense and should prove useful to any interested parties. It is difficult to comment on them effectively to suit all interests since these will vary from person to person. However, even a rapid glance over the computer generated tables can quickly assist in giving the reader a general impression of the digital data. A more detailed study enables an even greater understanding of the database characteristics. These quantitative results could never have been possible with similar orders of accuracy from the data in their graphic form.

1:2500 Scale

(a) Individual Map Sheet Analysis

The adopted sheet as one being typical of all the 1:2500 scale plans held at Durham was SU 6709, an area near Southampton. These results reflect the much smaller size of the 1:2500 data. The sheet results, which can be referred to in part a of Annex 4, comprises 136 features in all (111 are of line type and the remaining 25 are points). The entire data are held as 2,351 DMC records which is only seven per cent of the typical 1:1250 sheet for the Durham area.

On the whole, the detailed feature breakdown shows a similar picture to that depicted in the 1:1250 results. There are less feature types within the 1:2500 data although, in general, the mean segment length at ground scale is of a similar order. As in the 1:1250 data, feature code 30 is well represented accounting for nearly 6 km of the 12km total line length for the sheet at ground scale. This reflects the abundance of fences in a rural area which accounts for much of the linework of the large scale plans.

The mean segment lengths, on the whole, are of the order of 5 - 8 metres at ground scale (2 - 3 mm at plot scale) which is fairly acceptable for cartographic purposes. The longest segment in the entire sheet has a value of 186.5 ground metres being used to represent part of a river (feature code 59 - bank of double river). The minimum segment distance appears to be 0.06 ground metres.

The skewness and kurtosis values seem to have much more meaning in the rural areas, with the much more natural topography where straight line features are far less evident than in an urban environment.

(b) The Entire 1:2500 Scale Sheets Held At Durham

A total of 210 digital plans are held on magnetic tape at 1:2500 scale. Part b of Annex 4 looks at the entire 1:2500 data to discover that although there are fewer quantities, a similar trend to that discovered for the 1:1250 information does exist. Feature code 30 (fence, wall, etc. - non road) is still widely used being the most frequent of the 109 types of feature available. The generated line distance for code 30 accounts for an excess of 2,230 km (ground scale) which is far greater proportion--ally than in the 1:1250 data.

There are no indications of zero minimum segment lengths in the 1:2500 data although a value of 0.06m (at ground scale) seems to be very prominent. This figure may reflect the minimum resolution step possible on the Freescan digitisng tables in use at Southampton, or it could be a function of the software adopted by OS to generate further data points when the operators have underestimated the number required to fix a line with sufficient accuracy.

Of further interest with this scale data seems to be the incompatibility between the number of points used to fix the position of text strings and the associated frequency of occurrence of the text feature code (28). In the DMC format,. text position is fixed by a single point at the lower left hand corner of the first character in the string. Thus the number of points and the frequency for code 28 should be equal although this equality seems not to be present. According to the 'status' table, the 1:2500 data held at Durham possesses 11,114 text features although these only involve 11,104 points. It would follow that there must be ten instances where a feature is coded for text but cannot possibly contain any characters since no position for the text is supplied.

A second inconsistency discovered in the 1:2500 data is one taking a similar fashion to that outlined for the 1:1250 results. There were three occurrences discovered where the end of feature (-5) sequence codes have been excluded. The consequences of this have already been discussed.

The 210 sheets held at 1:2500 scale represents a ground area of 210 square kilometres and within this area, there are 5,571 km of line features at true scale. Of this number, one third is classified by feature code 30. There exists 65,024 features in these sheets, 75 per cent of which are lines whilst the remainder are point symbols or text strings. The number of ignored codes is proportionately lower than the 1:1250 data with 446 features coded as zero which accounts for in excess of 30,000 DMC records. There are 109 feature codes available for the basic scales data. The 1:2500 scale sheets held at Durham make use of 73 of these catagories which is less than the 1:1250 information which incorporates 87 categories.

The following feature codes are not used in any of the 1:1250 or 1:2500 data held at Durham:

38 - Telephone Call Box (AA)
40 - Telephone Post / Pillar (GPO)
41 - Telephone Post / Pillar (AA)
42 - Telephone Post / Pillar (RAC)
43 - Telephone Line
46 - Aerial Ropeway
56 - Antiquity - Detail
63 - Bank of Moat
73 - Ferry
74 - Generalised Coastline
75 - Contour

77 - Boundary - London Borough
87 - CL Motorway Roundabout
105 - Rare Feature Code - solid
106 - Rare Feature Code - pecked
107 - Rare Feature Code - symbol

(c) Stratification

Part c of annex 4 shows the statistical breakdown for the six strata in the 1:2500 data. The factors governing the choice of strata were kept as far as possible the same as those adopted for the 1:1250 data. It is, of course, difficult to find 1:2500 mapping in areas of large urban density since in these areas the 1:2500 scale usually gives way to the more detailed 1:1250 plans. However, West Sussex was considered to possess fairly high density urban properties and this factor reflects in the greater file sizes in these areas when compared with Cornwall, for example, which must be looked upon as having a fairly light urban density.

The stratum classified as 'Light Dense Urban / Undulating (Cornwall)' appears to have some anomaly since its size far outweighs the others by as many as 5000 records per sheet. However, this could be accounted for by the much reduced maximum segment lengths detected in the majority of the line features. This will produce a greater number of digitised points and hence the number of records required must rise accordingly.

On the whole, however, there appear to be fairly unsubstantial differences between the strata with perhaps the occasional anomaly or extreme case. On average most sheets seem to have the order of 200 - 300 features which entails a total line distance of around 12 metres at plot scale. These figures being much less than those for the 1:1250 data which have the order of 2000 features per sheet and a total line distance of circa 35m at plot scale.

Comparison Of The Digital Sheets At 1:1250 And 1:2500

Evidence has already been supplied to suggest that in all circumstances the 1:2500 sheets are much smaller (by a factor of 4 or 5 times) than the 1:1250 sheets. This difference must be a direct reflection on the great abundance of information which the latter scale possesses. This great variation in size reflects on all forms of computer manipulation. The 1:1250 plans require greater storage areas (circa 16 feet of magnetic tape for a sheet of 25,000 DMC records packed at 1600 bpi phase encoded on a 9 track tape drive). This increase in the storage area produces slower process time per sheet and a corresponding slower plotting time since there is more information involved.

There appears to be no substantial differences in the line segment distances between the two scales. The software presents this information at ground scale to aid comparison and although the total line lengths do differ quite considerably, the mean segment distances seem to tally (of the order of 5 - 7 metres at ground scale for most line features).

Although there are fewer features in a 1:2500 sheet, there is evidence to suggest that they require more DMC records proportionately than are required to store features of a similar nature at 1:1250 scale. This will no doubt stem from the reduced maximum segment length, the vast extents of rural linework and the differing characteristics of a rural area when compared with an urban region.

Best comparisons between the two scales could be made in an area where there is coverage at both scales. This type of area is scarce in the UK; the scale for basic mapping as outlined by Harley (1975) depends on the population figures it possesses. Within the digital sheets held at Durham, there is only one area with map coverage at both scales - an area in southern England represented by SU 6709 at 1:2500 scale and SU 6709SE and SU 6709 NE at 1:1250 scale. There is not complete overlap therefore, although the eastern two quadrants have coverage at both scales.

Annex 5 shows the results of this multi scale coverage. The output has been treated as a 'stratification' run by the program to allow easy comparisons. Even though the 1:2500 sheet represents a larger ground area than does the two 1:1250 quadrants, it can be seen how the smaller scale representation has been generalised since it only involves 21 feature types whereas the large scale coverage involves thirty. Feature code 30 seems to be very typical in most sheets and this comparison exercise shows quite conclusively how there are proportionately more points (and hence more line segments) in a single 1:1250 quadrant then there are in the equivalent 1km x 1km 1:2500 coverage for the area.

The number of text characters included is greater in the larger scale data with some 522 characters per quadrant as opposed to 125 for the whole of the 1:2500 extract.

On the whole, therefore, these statistical results give quantitative evidence to many facts which could be looked upon as merely assumptions before the digital era began. In conclusion, the 1:1250 coverage has more detail and involves more costs in terms of computer access time and storage whereas the 1:2500 data are much more generalised for any associated cartographic implications. Even though there are these fundamental differences in the basic scales, their characteristics tend to be very similar. There is no evidence in the data held at Durham to suggest that the accuracy at ground scale is any better for either scale; indeed results tend to indicate that each segment of a line feature (i.e. the ground distance between two digitised points) is manifestly the same at both scales.

Assessment Of The Results And Projection Of The Figures For The Entire Databank On Its Completion

Since the data sample held by Durham was subdivided into six strata, the breakdown of which is outlined in Appendix E, then these subdivisions could be used for a more accurate projection of results for the country as a whole.

There appears to be no evidence in the results so far discussed to indicate that the nature of the terrain has any influence on the digital characteristics of the information. Consequently, only population size within each strata was used as a governing factor for the projection. The sample was grouped together into three strata based on light, average and dense urban areas. In a similar manner, the country as a whole required to be subdivided into areas of similar category in order that an assessment could be made on the number of sheets located within each stratum. It is of paramount importance to ensure that this subdivision of the country was as accurate as possible although the task itself was to be extremely problemmatic.

The exercise was accomplished by coordinating the OS sample with the 1971 UK Census data for 1km grid square units supplied to the Census Research Unit of the University by OPCS (Office of Population Censuses and Surveys). Harley (1975) discusses the flexible rules governing the basic scale at which an area is to be mapped and applying these as a rough guide line, the country could be subdivided into six strata (three at 1:1250 and three at 1:2500) whereby the number of sheets in each could be obtained by a graphical approach.

Appendix F shows the results of a computer program written to plot a graph of Population versus Cumulative basic grid square for the entire country. A logarithmic y scale was adopted to produce a more suitable curve for extrapolation. In order to speed up the CPU time required, only 1km grid squares with 25 people or over were included as it was assumed that other areas would be mapped with a basic scale of 1:10000 and 1:2500.

The population values for the grid squares in which Durham has OS digital coverage were retrieved from the CRU databank. A scatterplot was included on the graph to produce a fairly substantial breakdown of the sample into the six strata required. Results showed that the flexible rules discussed by Harley (1975) for which basic scale to adopt in an area had been closely adhered to by Ordnance Survey.

Having obtained a stratification for the country to match those classifications adopted in the sample, it was then a fairly trivial task to project figures for approx. how many sheets from the country as a whole would exist in each of the six strata. The results were governed by two constraints: the number of sheets in the entire large scale series and the numbers in both the 1:1250 and 1:2500 series.

Table 4.2 below shows the estimated number of sheets existing in each stratum for the entire country. Also given are the estimated mean population values per sheet.

Table 4.2 - Projected Stratification Breakdown For The Country

1:1250 Scale	Light	Average	Dense	Total
No. of sheets	20708	17012	15596	53316
Mean pop. per sheet	1437	2366	4930	
1:2500 Scale				
No. of sheets	89202	52070	23189	164461
Mean pop. per sheet	15	68	296	

Having determined the expected number of sheets in each stratum, the sample statistics were then projected to their equivalent values for the entire country. Primarily, projections were made on the stratification basis and then these were grouped together to produce a single set of estimate results for the whole database on its completion. These results can be referred to in Appendix G.

Assessment Of The Projected Results

Any projected figures at such an early stage of data availability must be considered with caution. The introductory text of this thesis has attempted to suggest that these results provide an exploratory tool only for any subsequent, more detailed, studies.

The projected figures as outlined in Appendix G can be no better than the size and structure of the basic sample. However, attempts have been made to keep the approach to as high a standard as is possible given the underlying constraints of data availability, previous research of a similar nature and time.

The results suggest that the entire data structure (1:1250 and 1:2500) will require of the order of 18,000 megabytes of storage or 581 magnetic tapes (stored in DMC as efficiently as possible - 1600 bpi, 2400 feet, 9 track). It is expected that there will be in excess of 47×10^6 coordinate pairs in the entire databank.

The figures included in Appendix G are very large as should be expected. Also supplied, with some reluctance, are margins of error for the estimated values. These margins of error are given for the 95 per cent confidence limits although they should be used with care since the moment statistics suggest that in many instances, the distributions are not near normal. These margins are of a similar high order which suggests that caution should be placed on any of the projected figures. Earlier results have already shown that the segment length distributions are not Gaussian in nature and hence any confidence limits included should carry little importance. However the absolute counts have a fairly high order of normality and in this way moment statistics suggest that the confidence limits have more meaning.

The information given for the mean, maximum and minimum segment distances cannot be projected and can only be suggested on the basis of the figures obtained from the sample.

There are several instances where specific feature types are not included in the results. This situation arises as a result of their absence from the sample data. Consequently no projections can be given for these feature codes in the completed databank, although their existence is more than likely to be evident.

The 1:2500 data will require more storage and computer processing time than will the 1:1250 data. This situation arises because there is an excess of 111,000 more 1:2500 sheets than there are 1:1250 sheets. Consequently the total number of lines, points, etc., are correspondingly higher in the 1:2500 section of the databank. It is interesting to notice, however, that projections suggest that the 1:1250 section will have of the order of 20×10^6 more features than have the 1:2500 data.

Finally, projected trends suggest that some 37×10^6 ignored DMC records will exist which is equivalent to ten computer compatible tapes. It would follow that great emphasis should be placed on attempts to ensure that these minor inconsistencies are kept to a minimum.

Chapter Five

The Feasibility Of A 1:50000 Digital Topographic Database

It was in the period between 1974 and 1976 when the seventh series of the one inch to the mile sheets were replaced by the metric 1:50000 scale series. (Harley 1975). Each graphic within the 800mm x 800mm neat lines represents a ground area of 40km x 40km. There are some 204 of the new sheets providing complete coverage of the UK. This is an increase of fifteen sheets over the preceeding 1:63360 series. The use of six colours in the graphics is perhaps the most immediate noticeable difference over the basic scales data although, of course, at such a reduced scale the process of information generalisation has the most fundamental implications from a topographic database point of view.

The basic scales information at 1:1250 and 1:2500 is used quite successfully by Ordnance Survey for the derived mapping series of 1:10000 scale (when this itself is not the basic scale) and at 1:25000 scale (Searle and Waters 1979). Such an approach requires a certain level of information selection but certainly elegant results can be produced at both derived scales as Searle and Waters (1979) show, using the basic scales information as the source data. In the production environment at OS, the Ferranti Master plotters are producing highly detailed plots with extreme accuracy which was often not possible by manual draughting techniques.

The design of the 1:50000 series involves far more cartographic generalisation than is required for large scale mapping and involves the use of symbology far more for point and linear representations. Harley (1975) points out that the choice of symbols was based on the results of map perception studies undertaken in the early seventies. In addition, this series includes height information in the form of point altitudes and contours. The contour interval on sheets in the second series is 15.24m - based on the traditional 50 feet C.I. of the one inch series.

Drewitt (1973) points out that market research undertaken by OS has suggested that the largest single group of users for 1:50000 information is likely to be motorists. As a consequence of this, great emphasis has been given in this mapping series to the depiction of routeways and communication links. The great significance given to this type of information, together with boundaries and water features has tended to produce maps comprising of many linear representations. These factors have direct and important implications on the conversion of a 1:50000 sheet into its digital equivalent.

It has been discussed elsewhere (Thompson 1978) how various agencies have become actively involved in the production of pilot digital mapping schemes for 1:50000 scales or smaller. Ordnance Survey themselves have recently commenced feasibility studies into the possibilities. The implications of such a service are widespread and may be better suited to some applications which are at present utilising the large scales. Studies associated with land use, ecology, transport networks and geology immediately spring to mind.

The preliminary tests carried out in house at Southampton has involved a study of sheet 202 (Torbay and South Dartmoor). These tests have entailed similar digitising and edit phases as undertaken in the large scale production flowline. At the early stage of this M.Sc. project, three basic 1:50000 feature types had successfully completed each phase. These data were representations of water features, the centre lines of roads and boundary information. Together with the sample of the large scale data, a magnetic tape was supplied to Durham providing a copy of these small scale data. Since this information was written in the DMC customer format, a similar statistical analysis could be carried out .

It has been the policy of OS, during this feasibility investigation, to separate each feature type into its own individual file, thereby treating each type of feature as a separate sheet. Concatenation of features within the D09 plotting program is still possible, of course, in the usual way.

The water features information which includes such features as rivers, streams, canals, HWM, swamp, drainage area boundaries, etc., was digitised at 1:55000 scale and then digitally enlarged to correct scale. As results show, the sheer magnitude of this information in its digital format is a major disadvantage. As a consequence, the roads and boundary data were separated into quadrants and thus treated as eight separate map sheets (four for each feature). These data were digitised at 1:30000 and digitally reduced to correct scale.

Statistical Analysis Of The 1:50000 Digital Data

The results of the statistical analyses of the 1:50000 digital data held at Durham can be referred to in Annex 6. The vast number of points required to digitally store the water information for sheet 202, which is essentially a series of linear representations, becomes quickly evident. In excess of 85,500 x and y coordinates are required to digitally represent these data. This vast number of coordinate pairs reflects in the number of DMC records required in total with over 108,000 being involved in the storage of the water information alone. Although there are probably no more than 1500 linear features within the water data, there are in fact some 82,244 line segments involved for the accurate regeneration of these features. On average, the mean ground distance delineated by these segments seems to be of the order of 8 - 9 metres (0.17mm at plotting scale). This is obviously highly accurate given that the information has been acquired by manual digitising techniques.

An anomaly was detected for feature code 201 (HWM - estuary mouth). This feature boasts a mean segment length of 841 metres at ground scale. The feature, which has a frequency of occurrence only four, appears on the graphic as a series of four straight lines crossing the Kingsbridge estuary near Salcombe, South Devon. This information, together with others such as the point position of water sources and mouths, was included in the 1:50000 data files for the benefit of the water development unit.

There is evidence from the statistical analyses to suggest that the 1:50000 data have several zero length minimum line segment distances as was also detected in the 1:1250 data. The reasons for the presence of this inconsistency (or inefficiency) in the data have already been discussed. No doubt the reasons for its presence here are based on similar arguments.

As was to be expected, the total ground distance represented by all the linear features is fairly large. The program has reported a total line distance at ground scale in excess of 800km (1,600cm at 1:50000) for the water features alone. Of this total distance, some 327,923 ground metres are classified as single streams (feature code 64). This type of information could never have been acquired with similar accuracies by conventional means. The introduction of digital mapping techniques has enabled quantitative approaches to be possible in so many fields with obtainable accuracies of which one can be certain.

The skew and kurtosis values tend to be very large and unpredictable in both the small and large scale data sets. This is probably to be expected in the large scale information since the majority of line segments bear little relationship to each other. In an urban environment there is little evidence of many curved features requiring a series of short length connected line segments for their representation. On the whole the topography consists of many longer length line segments which are uncorrelated.

However, with the small scale data, especially within features such as rivers, coastlines and boundaries, the existence of long straight lines is not to be expected. This hypothesis is given further evidence with the short maximum segment lengths which have been detected. Since the skew and kurtosis statistics are not zero or near to zero, this suggests that the distribution of the line segment lengths is not Gaussian (normal). Quantitative geographers such as J. B. Kruskal, J. W. Tukey, I. S. Evans and others have maintained that to re-express (transform) the original data is trivial, especially by computer. By making the distribution nearer to normality, it becomes possible to use the benefits of parametric statistical techniques.

There is evidence, from the large positive skew and kurtosis values, to suggest that the distribution has a strong positive skew. A trivial transformation experiment involving taking the logarithms (base 10) of each line segment produced the desired effects in making the distributions nearer to normality. Included in Annex 6 part c are the obtained measures for skew and kurtosis after the transformations. Since they are nearer to zero, then it is possible to conclude that the segment distributions in their untransformed state have a strong positive skew. On the whole, however, the segment lengths are small with few larger anomalies. Thus it has been possible to give quantitative evidence for the characteristics which are to have been expected in the first place .

The boundary and centre lines of roads information appears to have very similar characteristics although these files have smaller absolute counts since they are divided into quadrants. The quality of the information is similar to the figures derived for the water features.

After digitally concatenating the four quadrants ~~together~~, the following counts were obtained. Some 42651 data points involving 61,757 DMC records were required for the road data and 22511 data points involving 26,431

DMC records were required for the boundary data. The mean segment lengths appear to be larger on the whole than the figures obtained for the water data, with an average ground distance of 50m (1mm at 1:50000). The skewness and kurtosis measures suggest near normality even before transformation, although one based on logarithms does improve the situation even further.

The results report some 1,530km of roadways on sheet 202, the bulk of which is classified by code 97 (minor single carriageway less than 4m). There are approximately 2,700 features included in the roads data for the entire sheet. The boundary information comprises of over 22,000 line segments, 500 features and a total ground distance of 717km (14.3m at 1:50000). One can be sure that this information is correct although the accuracy and characteristics of the data themselves remain totally in the control of the digitising operators.

In conclusion, there is evidence to suggest that the characteristics of the small scale data do differ from those of the larger scale. Naturally, the 1:50000 data acquisition is still in its early stages although even now there is evidence to suggest that the storage requirements and processing time involved will be quite substantial. As a rough guide, the three categories of 1:50000 information held at Durham requires of the order of 130 feet of magnetic tape (9 track, 1600 bpi, phase encoded, DMC) and a computer CPU time in excess of 100 seconds purely for I/O. It should be kept in mind in addition to these figures, that a 1:50000 sheet contains many more than three categories of feature. Some of these will be very large consumers of backing store. Contour data are a typical example. Since there are 204 sheets in the entire series, then these figures will swell to alarming proportions.

The Benefits Of A Scale Independent Database

Although the National Mapping Agency is currently involved in investigations leading to a 'structured' Land Information System, much of the preceeding developments have been based on the cartographic applications of the topographic data. The investigations into the production of a digital databank for 1:50000 information is an extension of the cartographic criteria. Present OS policy regarding the 1:50000 database, although of course development is still at a preliminary stage, is to keep the large scale and small scale data sets independent. Possible links could be developed in the long term future should they be required.

Many could argue that this policy is somewhat restrictive and unnecessary. Indeed some critics suggest there to be no useful requirement for 1:50000 data at all. The concept of topographic data scale suggests the over emphasis and importance being placed on any cartographic implications which may exist. The notion that a scale free Land Information System would be the ideal solution is held by many interested parties. The United Kingdom is fortunate in being one of the best mapped areas of the world with the largest scale providing complete ground coverage being 1:10000. In order to ensure data homogeneity in terms of accuracy, 1:10000 would have to be the scale for the database. However, since many regions of the country are mapped at larger scales, then the inability to include this information into the system would be a major disadvantage.

The cartographic implications should not be completely neglected in the concept of a scale free database. The basic scales 1:1250 and 1:2500 information have a higher order of accuracy than 1:50000, say, and is more advantageous for computer storage. Digital techniques could then enable the appropriate selection and generalisation for considerations at smaller scales. The 1:50000 data, on the other hand, are not completely mere generalisations of the larger scales. There are instances where the small scale information includes features which are not represented on a large scale plan. The classification of tourist viewpoints, historic sites and beauty spots are typical examples.

To provide a useful and working system, therefore, it is impossible to completely ignore the effects of scale. Although present technology restricts the concept of scale independence becoming a reality, there is no reason why attempts should not be made to develop a variable scale database. This notion would involve digitising all areas of the country at some varying feasible scale (being dependent on the information available). Links could then be provided between the areas to enable cross referencing and manipulation within the bounds of the data accuracy. This 'patchwork quilt' approach could then be gradually improved to some uniform large scale as technology and hardware becomes more efficient and cheaper. It is quite reasonable to suggest, therefore, that a small scale digital database becoming commercially available in the short term future would provide a useful intermediary between the present situation and the ideal solution which, as Thompson (1978) suggests, will only become available in the long term future.

Chapter Six

Conclusions

The advantages of supplying topographic information in a digital format are widespread offering far more versatility than has hitherto been available. The distribution of material from a continuously up-dated digital topographic database offers the capability of providing 'customised images' with features, area of interest and scale being selected by the user.

The major drawbacks of the current production system involve the form in which the data are supplied. Attempts have been made in this thesis to highlight these limitations from two points of view: the need for structure in the data and problems of inefficient input and output with a data processing installation.

The need for data reorganisation dates from the early days of OS involvement in digital techniques. Following initial feasibility studies, a development project to produce software to restructure the OS digital data was commissioned from PMA Consultants (Thompson 1978). In mid 1979, this project is nearing completion and is designed to enable the use of the data for a wide range of user needs, both graphical and analytical, avoiding duplication in the production and storage of the spatial information (Thompson 1979). The development of this restructuring software was logical and necessary and, as a consequence, the market of OS digital data is expected to expand.

The problems of slow I/O with the DMC format arises, especially in a multiple access computing environment, as a result of the card image nature of the data. The character conversions and data validations are slow and costly in terms of computer resources; certainly the present sequential searches required for specific map sheets by the plotting software is unreasonable. In order to alleviate the problems it becomes necessary to reorganise the data format into a binary system allowing large blocks of information to be moved around the computer installation by means of executive I/O, unencumbered by time consuming data conversions. The adoption of these ideas will provide an interim solution although it still remains necessary to design an efficient database manage--ment system involving addressing cross references and index files to ensue the inefficient sequential searches.

The Ordnance Survey cannot be expected to resolve these problems entirely since much of the data addressing and storage in backing store is machine dependent and until an interface at binary level is designed between OS hardware and all customer machines then the utilisation of data transfer in character code must remain. In the long term, however, the rates of change of computer-based technology are such that alternative approaches of data interchange will be possible. The digital transmission of material over a national data communications network would obviate the need for user-based storage completely.

The improvements in technology should be beneficial to system efficiency from a data acquisition point of view. Thompson (1978) concludes that an important issue to be resolved sooner rather than later is a reduction in digital mapping costs. The adoption of second generation digitising techniques involving data capture and on-line editing could provide more efficient solutions.

The question of data scale must remain controversial for the present time. Certainly the completion of the large scale database must be considered in the long term only. The adoption of a small scale interim database has much potential although feasibility studies at Southampton and Durham suggest that the problems of data volume could well prove too problematic for the current handling techniques. The limitations of slow I/O and information retrieval increases proportionally with file size and since exploratory studies have shown the high volume of data for each 1:50000 sheet, then a more improved storage and referencing system must be made available.

The future must see the introduction of an improved system possessing the enhancements which have been discovered during the past four years as a result of developing the DMB system. This second generation system must adopt the most up-to-date techniques available in both hardware and software terms to ensure that the data content and accuracy remains consistent. In this way the data can not only be used for graphical applications but also the most rigorous of analytical tasks. The statistical analyses provided by this project suggest that the data accuracies are sufficient for most purposes although the lack of post processing routines could affect certain tasks. The inability of property boundaries to constitute closed polygons was an example cited in the text.

On the whole there are certain trends in all large scale digital sheets (at 1:1250 and 1:2500) irrespective of the areas which they represent although urban population is strongly correlated with data volume. Ground terrain has no affect on data accuracy or content at the basic scales. An interesting theme arising from this thesis involves correlation studies between OS digital information and the census data. Since the University at Durham is fortunate to hold samples of both these data sets, it is the intention of the author to investigate these ideas further in the near future. It should be possible, for example, to attempt correlation studies between population and total line length of property boundaries to discover the strengths of any correlation which may exist.

In conclusion, this work has provided an introductory example of manipulation, other than graphical, to which digital topographic information can be put. In so doing, it has been possible to discover limitations in both the approach adopted for the study and the digital service currently being offered by Ordnance Survey. It should be pointed out that the Mapping Agency is aware of many of these drawbacks and the development branch is continually striving to improve the service. It is anticipated that in the early 1980's a prototype second generation system will be installed at Southampton. This new system, currently being referred to as the Topographic Database Production System (TDPS), is expected to run parrallel with DMB for an initial period. However, in the longer term it is hoped that many of the drawbacks of DMB will be overcome, namely cost, speed and data structure.

The gradual improvement of the digital service and the widespread applications to which the information can be put will only gain importance if the users, with their inbuilt resistance to change, can be successfully converted to appreciate the benefits which this new technology can provide.

Acknowledgements

I am indebted to many friends and colleagues whose encouragement and ideas have stimulated me into investigating many approaches to the problems encountered.

I particularly wish to record my gratitude to Mahes Visvalingam, Ian Evans, Nicholas Cox and Ray Harris of the University of Durham who have continually guided and assisted me in all aspects of data analysis and computer applications.

Many thanks are expressed to Col C. N. Thompson, Capt P. R. Searle and Mr R. S. Waters of the Development Branch at Ordnance Survey who have provided the data and offered much advice and encouragement on technical matters.

Finally, I am especially indebted to David Rhind of the University of Durham who has inspired and supervised the entire project, expressing many opinions and ideas providing the initial stimulus for the interests and experiences now held by the author.

T.A.A. September 1979

Bibliography

BAXTER R S (1976) — Computer and statistical techniques for planners. Methuen

BERRY B J and MARBLE D F (1968) — Spatial Analysis - a reader in statistical geography. Prentice-Hall

DALE P F (1977) — The land surveyor and land information management. FIG Conference Sweden 1977

DREWITT B (1973) — The changing profile of the map users in Great Britain. Cartographic Journal 10

GARDINER-HILL R C (1972) — The development of digital maps. OS professional paper no. 23

HAGGETT P, CLIFF A D and FREY A (1977) — Locational Analysis in human geography volume 2. Arnold

HARLEY J B (1975) — Ordnance Survey maps - a descriptive manual. HMSO

HARRISON J C (1978) — The LAMIS system. RICS/NELP conference on databanks and digital mapping Nov 1978

MARLES A C (1971) — The place of map miniaturisation in map production. Conference of commonwealth surveyors Cambridge 1971

Ordnance Survey — D09 FORTRAN plotting program description

Ordnance Survey	Digital data on magnetic tape
RHIND D W (1973)	An introduction to the digitising and editing of mapped data Annual symposium of the British Cartographic Society, 1973 BCS special publication no. 1
RHIND D W (1974)	The state of the art of geographic data processing - a UK view Seminar of geographic data processing, Peterlee 1974
RHIND D W (1978)	The storage and retrieval of land related data RICS Annual conference, Cambridge
RHIND D W and TREWMAN T (1975)	Automatic cartographic and urban databanks - some lessons from the UK International yearbook of cartography 1975
SEARLE P R and WATERS R S (1979)	The Ordnance Survey 1:10000 and 1:25000 digital maps Conference of commonwealth surveyors, Cambridge 1979
SIMMONDS M C (1978)	The presentation of survey information on microfilm Institute of reprographic technology/OS 1978
THOMPSON C N (1978)	Digital mapping in the Ordnance Survey 1968 - 1978 RICS/NELP Conference on databanks and digital mapping Nov 1978
THOMPSON C N (1979)	The need for a large scale topographic database Conference of commonwealth surveyors, Cambridge 1979

VISVALINGAM M, NORMAN M J and SHEEHAN R (1976) Data interchange on industry compatible tapes Census research working paper no. 6 University of Durham

Appendices

The annexures to this thesis can be referred to in an accompanying line printer formatted file ɔDNANCE SURVEY DIGITAL DATA

ɔrmats on Fortran readable magnetic tape

DESCRIPTION	CONTENTS CHARS 1-4	CONTENTS CHARS 5-8	FORTRAN FORMAT
tart sheet:	-1	0	2I4
SW corner co-ords (NG) (E		Reference - E	)
(N		Reference - N	) 4F8.0
Size of basic grid square (metres)		Size of square	)
Source scale (reciprocal)		Source scale	)
ıd of sheet	-2	0	2I4
ıd of file	-3	0	2I4
art of feature	-4	Serial No	2I4
d of feature	-5	0	2I4
ature type (line)	-7	Feature code	2I4
ature type (text, symbol)	-8	Feature code	2I4
sic grid square indicator	-9	0	2I4
B.G. square reference E,N	Ref E	Ref N	2F4.0
xt indicator, classification	-10	classification	2I4
Style, height of characters	Style	height (mm)	I4, F4.1
xt character count	-11	count	2I4
Characters	4 characters	4 characters	2A4
ientation (all features)	-12	Angle (units of .1 degree)	2I4
ag (invisible line)	-13	0	2I4
tance	-15	distance	2I4
tour value	-17	height (metres)	2I4
nt co-ordinates curate to .001 of a BG square)	E Value	N Value	2F4.3
action (7 track only)	- 6	0	2I4

Appendix B

Topographic Feature Codes Used In DMC

Codes	Type	Description
1	Line	Building - Public
2	Line	Building - Minor
3	Line	Building - Other
4	Line	Building - Open sided
5	Line	Building Division - Broken line
6	Line	Archway symbol
7	Line	Boundary - Parish or Community
8	Line	Boundary - District
9	Line	Boundary - County or Region
10	Line	Boundary - Electoral division
11	Point	Boundary - Post or Stone
12	Point	Boundary - Mereing Symbol Full
13	Point	Boundary - Mereing Symbol Half
14	Line	Railway - Narrow guage
15	Line	Railway - Standard guage
16	Line	Railway - Underground
17	Line	Boundary - general
18	Line	Railway - Disused, Centre line
19	Line	Railway - Dismantled Centre line
20	Line	Railway - Switch
21	Line	Road pecks (carriageway)
22	Line	Road - Centre line
23	Line	Path (um)
24	Point	Minor control point
25	Point	Trig. point
26	Point	Bench mark
27	Point	Surface level
28	Point	Name/Number, position (text)
29	Line	Road fence, wall etc (casing definitive)
30	Line	Fence, wall etc - Non road
31	Line	Road pecks (casing definitive)
32	Line	Surveyed Pecks (banks, baulks, made paths, driveways, etc)
33	Line	Tunnel alignment
34	Line	Subway/Underpass alignment

Codes	Type	Description
35	Line	Vegetation limits (sketched pecks)
36	Line	Ground surface feature limits (sketched pecks suppressed)
37	Point	Telephone Call Box - GPO
38	Point	Telephone Call Box - AA
39	Point	Telephone Call Box - RAC
40	Point	Telephone Post/Pillar - GPO
41	Point	Telephone Post/Pillar - AA
42	Point	Telephone Post/Pillar - RAC
43	Line	Telephone Line
44	Line	Pipe Line - Obstacle
45	Line	Pipe Line - Suspended or Non Obstacle
46	Line	Aerial Ropeway
47	Line	Electricity Transmission Line
48	Point	Electricity Pylon - Standard
49	Point	Electricity Pylon - Surveyed
50	Point	Electricity Pylon - Part
51	Point	Electricity Posts (Surveyed)
52	Line	Step treads
53	Point	Cave Symbol
54	Point	Antiquity Symbol
55	Line	Antiquity pecks (course of)
56	Line	Antiquity - detail
57	Point	Point features - dot
58	Point	Objects shown by circle (non water 0.60mm)
59	Line	Bank of double River/Stream
60	Line	Bank of canal
61	Line	Bank of Double Drain
62	Line	Bank of Lake/Pond
63	Line	Bank of Moat
64	Line	Single Stream
65	Line	Single Drain
66	Line	Centre Line of Double Water Feature
67	Point	Flow arrow - large
68	Point	Flow arrow- medium
69	Point	Flow arrow - small
70	Point	Culvert bar
71	Line	Mean high water, MHWS
72	Line	Mean low water, MLWS

Codes	Type	Description
73	Line	Ferry
74	Line	Generalised coastline
75	Line	Contour
76		not used
77	Line	Boundary - London borough
78	Line	Boundary - Ward
79	Line	Boundary - Parly Const
80	Line	CL - Track
81	Line	Track
82	Point	Objects shown by circle (Water, 0.60mm)
83	Line	Road pecks (pavement etc)
84	Line	Ground surface feature limits (sketched pecks)
85	Line	CL Motorway Dual carriageway
86	Line	CL Motorway Single carriageway
87	Line	CL Motorway Roundabout
88	Line	CL Motorway Slip road
89	Line	CL Trunk/Main Dual carriageway
90	Line	CL Trunk/Main Single carriageway
91	Line	CL Trunk/Main Roundabout
92	Line	CL Secondary Dual carriageway
93	Line	CL Secondary Single carriageway
94	Line	CL Secondary Roundabout
95	Line	CL Minor Dual carriageway
96	Line	CL Minor single carriageway more 4m
97	Line	CL Minor single carriageway less 4m
98	Line	CL Minor - other roads
99	Line	CL Minor Roundabout
100	Line	CL Railway - multiple track
101	Line	CL Railway - single track
102	Line	CL Railway - sidings
103	Line	Alignment feature
104	Line	Normal tidal limit
105	Line	Rare feature code - solid
106	Line	Rare feature code - pecked
107	Point	Rare feature code - symbol
108	Line	Surveyed pecks (banks etc) type B
109	Line	Railway - buffers/retarders

Codes	Type	Description
110	Point	Campsite symbol (1:50000)
111	Point	Caravan site symbol
112	Point	Leader arrow
113	Line	Public right of way - bridleway
114	Line	Public right of way - footpath
115	Line	Public right of way - road used as path
116	Line	Forest park outline
117	Line	National Park/New forest outline
118	Line	National Trust boundary (always open)
119	Line	National Trust boundary (open restricted)
120	Line	Airport runway pecks
121	Line	Airport perimeter track
122	Point	Heliport
123		not used
124	Line	CL narrow guage railway
125	Line	CL private railway
126	Line	CL underground railway pecks
127	Line	CL railway under bridge
128	Point	Principal railway station
129	Point	Other railway station (in use)
130	Point	Disused railway station
131	Point	Mouth of railway station
132	Line	Railway viaduct or bridge
133	Line	CL Motorway under construction
134	Line	CL proposed motorway
135	Line	CL main road (single c'way) under constr.
136	Line	CL main road (dual c'way) under constr.
137	Line	CL narrow main road
138	Line	CL narrow secondary road
139	Line	CL underground road pecks
140	Line	Road under bridge
141	Line	Path/Towpath
142	Point	Road tunnel mouth
143	Line	Bridge/Subway/Viaduct
144	Point	Elevated road symbol
145	Point	Toll
146	Point	Footbridge
147	Point	Gradient arrows (greater than 1 in 5)
148	Point	Gradient arrows (greater than 1 in 7)
149	Point	Bus or coach station

Codes	Type	Description
150	Point	Motorway service area
151	Line	Building line (non road)
152	Line	Building line (road)
153	Line	Vegetation boundary fence (non road)
154	Line	Vegetation boundary fence (road)
155	Point	Standard roundabout
156	Line	Miscellaneous boundary (fenced)
157	Line	Miscellaneous boundary (unfenced)
158	Line	Ornamental park boundary (non road)
159	Line	Ornamental park boundary (road)
160	Line	conduit or exposed pipe
161	Line	Cutting
162	Line	Embankment
163	Line	Groyne
164	Line	Rifle range sightline
165	Line	Ski-lift/Chairlift
166	Line	Rope drag
167	Point	Landmark (circle)
168	Point	Church (with spire)
169	Point	Church (with tower)
170	Point	Church (with no spire or tower)
171	Point	Gas/Oil/Sewage tank
172	Line	Gate on road
173	Point	Golf course/Links
174	Point	Youth Hostel
175	Point	Mile post or stone
176	Point	TV/Wireless mast
177	Point	Windmill
178	Point	Windpump
179	Line	Roman road pecks
180	Point	Non-Roman antiquity
181	Point	Antiquity (not to scale)
182	Point	Battlefield
183	Point	Antiquity wall and other outline features
184	Point	Mile turret
185	Point	Trig pillar
186	Line	Bathymetric contour (normal)
187	Line	Bathymetric contour (accentuated)
188	Line	Bathymetric contour (obscurred)

Codes	Type	Description
189	Point	Height (levelled/unlevelled) (normal)
190	Point	Height (levelled/unlevelled) (near bdy.)
191	Line	Limit of cliff/scree
192	Line	Limit of marsh/salting/osier bed
193	Line	Limit of sand/mud/shingle
194	Point	Information centre
195	Point	Picnic site
196	Point	Parking area
197	Point	Viewpoint
198	Line	Bank of tidal estuary
199	Line	CL tidal estuary
200	Line	Tidal part of single stream
201	Line	HWM - Estuary mouth
202	Line	Swamp/Underground water alignment
203	Line	Drainage area boundary
204	Line	CL canal (wet)
205	Line	Drain up to 8m wide
206	Point	well
207	Line	CL canal (dry)
208	Line	Canal lock
209	Point	Isolated water feature
210	Point	Mouth of water feature
211	Point	Source of water feature
212	Point	Start of drainage area
213	Point	End of drainage area
214	Line	Waterfall/Weir
215	Point	Coastal beacon
216	Point	Lighthouse (used)
217	Point	Lighthouse (disused)
218	Point	Single building (min size)
219		not used
220	Line	CL Trunk road (narrow)
221	Line	CL Trunk road (single carriageway)
222	Line	CL Trunk road (dual carriageway)
223	Line	Limit of flat rock
224	Line	Distribution water main
225	Line	Trunk water main
226	Line	Domestic supply water main
227	Line	Abandoned water main (distribution)

Codes	Type	Description
228	Line	Abandoned water main (trunk)
229	Line	Abandoned water main (domestic)
230	Line	Zone boundary
231	Point	Air valve
232	Point	Valve
233	Point	Closed valve
234	Point	Pressure reducing valve
235	Point	Reflux valve
236	Point	Differential pressure point
237	Point	Hydrant
238	Point	Meter
239	Point	Capped end
240	Point	Water pumping station
241	Point	Water booster station
242	Point	Service reservoir
243	Point	Water tower
244	Point	Well or borehole
245	Point	Water treatment works
246	Line	Foul private sewer
247	Line	Surface water private sewer
248	Line	Combined private sewer
249	Line	Foul public sewer
250	Line	Surface water public sewer
251	Line	Combined public sewer
252	Line	Foul main sewer
253	Line	Surface water main sewer
254	Line	Combined main sewer
255	Line	Abandoned private sewer
256	Line	Abandoned public sewer
257	Line	Abandoned main sewer
258	Line	Direction of flow arrow
259	Point	Manhole (foul and combined)
260	Point	Manhole (surface water)
261	Point	Backdrop (foul and combined)
262	Point	Lamphole
263	Point	Sewage pumping station
264	Line	Miscellaneous line feature (thin)
265	Line	Miscellaneous line feature (medium)
266	Line	Miscellaneous line feature (broad)

Codes	Type	Description
267	Point	Miscellaneous circle feature (0.7mm dia)
268	Point	Air photo control point
269	Line	Archaeological site - line feature
270	Line	Archaeological site - outline of solid area
271	Point	Archaeological site - point feature

correct in May 1978

ORDNANCE SURVEY DIGITAL DATA

Sample contents of a Fortran readable tape (file) containing one 1:1250 map of four features.

	Record contents							
	1	2	3	4	5	6	7	8
Start sheet			-	1				0
			4	3	7	5	0	0
			1	7	8	0	0	0
							5	0
					1	2	5	0
Feature 1			-	4				1
			-	7				3
	-		-	9				0
				2				3
		3	5	0			5	5
		-	1	2		1	5	0
		4	3	0			7	5
		4	1	0		1	5	0
		3	3	0		1	3	0
			-	5				0
Feature 2			-	4				2
			-	8			2	8
			-	9				0
				3				4
		5	0	0		8	5	4
		-	1	2		6	0	0
		-	1	0				4
				3			1	8
		-	1	1			1	1
	H	I	G	H		S	T	R
	E	E	T					
			-	5				0

	Record contents							
	1	2	3	4	5	6	7	8
Feature 3			-	4				3
			-	7			3	0
			-	9				0
				2				1
		3	2	5		9	0	0
		-	1	2	2	4	5	0
		1	8	0		7	5	0
		2	0	5		9	5	5
			-	9				0
				2				2
		2	8	7			9	8
		-	1	3				0
		2	5	9		2	0	7
		3	0	8		3	5	0
			-	5				0
Feature 4			-	4				4
			-	7			4	9
			-	9				0
				5				1
		1	8	3				7
		-	1	2		9	0	0
		-	1	5			5	4
			-	5				0
End of map			-	2				0
End of file			-	3				0

ORDNANCE SURVEY PLOT PROGRAM D09

May 1978

ORDNANCE SURVEY PLOT PROGRAM

1 DESCRIPTION

Written in standard FORTRAN, the program has a number of options selected by the user through parameters input on cards or paper tape.

Map data is input from FORTRAN readable magnetic tape in character format.

Output from the program as supplied is in the form of basic plotting data placed in a specified 'common area' in the computer core store. The user will write* and compile routines into the program to manipulate the data as necessary and use it for calls to his plotter software. Each entry in the common area supplies data for one call to a line, symbol or text generation routine for the plotter.

The program can be controlled to output data for one or more separate plots, each composed of data from one or more maps on the input tape. The plot area will be rectangular, but need not coincide with the boundaries of the component maps.

The input tape may be rewound at any time to use or re-use maps stored nearer the start of the tape.

* Ordnance Survey can supply suitable routines for some plotters.

1.1 SPECIFICATION

Program identity: D09

Source language: NCC Fortran

Peripherals: 2 magnetic tape decks, 1 paper tape or card reader, 1 line printer (optional)

Compiled size (1900): 15 000 words

Construction: Modular (14 modules)

Input:

M/T: A file containing data for one or more maps in fortran readable character format (see 'Ordnance Survey Digital data on Magnetic Tape')

Card or P/T: user control parameters, see para 2.

Output:

Common area in core: Basic plotting data in fixed locations, see para 3

L/P: Feature listing (optional), see para 2.3.4

Output for plotter: Controlled by user routines

Input/output channels allocated

1 Tape/card reader

2 Input magnetic tape

3 Line printer

2 CONTROL PARAMETERS SET UP BY USER

Parameter types, identified by numeric codes 01-23, may be input on paper tape or cards, giving the user a number of options when setting up the plot program.

A list of all parameter types with associated data is at 2.1, and the sequence of entry and effect of the parameters is at 2.2.

A plot may consist of one or more full sheets (code 01) or parts of one or more sheets (code 02), part sheet limits may be defined to an accuracy of 100 metres (1:2500 source) or 50 metres (1:1250 source).

One or more plots may be defined for a run, re-using the same sheets if required.

2.1 CONTROL PARAMETERS

CONTROL PARAMETERS

Type Code	Description	Essential/ Optional E or O	Parameters (Values are fixed length with leading spaces or zeroes)	Notes
01	Plot rectangular area (full sheets used)	E	) NG co-ordinates of SW and NE limits) of plot entered as four eight digit) values	Use of either 01 or 02 essential
02	Plot rectangular area (part sheets used)	E	))	No grid is available with type code 02
03	Feature description table (FDT)	E	See 2.3.1	
04	Plot scale	O	Integer value (eg 1:2500 = 2500) eight digits	If omitted, source scale of first required map is used
05	Text table	O	See 2.3.2	If omitted, text uses pen/ projector value in FDT, and size is that given in the feature.
06	Symbol table	O	See 2.3.3	If omitted, only standard plotter symbols can be output.
07	Report option	O	Nil	If omitted, no printer report is output.
08	Grid indicator	O	Grid interval (metres), eight digits	If omitted, a neat line is drawn round the plot area.
09	End of parameters	E	Nil	
10	Reject by feature code	O	Feature codes, four digits each, terminator, negative value	
11	Reject by name category	O	Category numbers, two digits each, terminator, negative value	See 2.3.5 for list.
12	Reject by serial number	O	Serial numbers, four digits each, terminator, negative value	

2.1 CONTROL PARAMETERS (Contd)

Type Code	Description	Essential/ Optional E or O	Parameters (Values are fixed length with leading spaces or zeroes)	Notes
13	Select by serial number	O	Serial numbers, four digits each, terminator, negative value	Overrides codes 10,11, but feature code must be in FDT
14	Sheet to which following parameters apply	O	NG co-ordinates of SW corner, two eight digit values	Following parameters apply when specified sheet is reached.
15	Title position and size (see 2.3.6 for title contents)	O	Reference of the origin of title block relative to plot origin (mms ±), height of text (1/10mm units), two eight digit values, one six digit value	If omitted, title origin is set 3.5cm below lower limit of plot and inset 1/10 of plot width from LH side. Height of type: 5mm
16	Marginal data	O	See 2.3.7	Separate entry for each line of text
17	Plot serial numbers	O	Nil	
18	Locate plot origin	O	Distance (mm±) from plotter origin, two eight digit values	If omitted, SW corner of plot at plotter origin
19	Rotate plot	O	Rotation (1/10 degrees) anti-clockwise from horizontal, six digits	
20	Terminate plot	E	Nil	
21	Rewind input M/T	O	Nil	Required before accessing an earlier map on the tape
22	Cancel report option (07)	O	Nil	
23	Cancel serial number plot option (17)	O	Nil	

2.2 PARAMETER SEQUENCE AND EFFECT

There is no fixed sequence for entering parameters (other than 09), tables (03,05,06) should generally be set up first, particularly if using options 10-13.

Code	
01) 02)	Must be present in first plot parameter set, repeated at the start of each new plot but may be omitted if subsequent plots cover the area specified in the first entry.
03	Must be present in first plot parameter set, will remain effective until replaced by another 03 entry.
04	Applies to current plot only.
05) 06)	If inserted, will apply to current plot and remain effective until replaced.
07	Applies to current plot and remains effective until cancelled.
08	Applies to current plot and remains effective until replaced.
09	Must be final parameter, and must not be preceded by 20.
10) 11)	Apply to current plot and remain effective until replaced.
12) 13)	Apply to one map only (preceded by code 14 entry unless first map on plot), both may be used. Re-specify for each map requiring either option.
14	Map reference on input M/T, holds following parameters until required map occurs.
15) 16)	Apply to current plot and remain effective until replaced.
17	Applies to current plot and remains effective until cancelled.
18) 19)	Apply to current plot and remain effective until replaced.
20	Applies to current plot, unnecessary if end of plot coincides with end of input M/T. It must not precede a 09 parameter.
21	Entered whenever a rewind is required.
22	Cancels report option (07) until re-set.
23	Cancels serial number plot option (17) until re-set.

2.3.1 FEATURE DESCRIPTION TABLE (Option code 03) (FDT)

Plotting details for each feature code required for output are entered, features with codes not entered in the table cannot be plotted.

Table format

Up to 150 feature codes may be entered, with values in the range 1-150. For each feature code, four fixed length parameters must be present.

An entry consists of 4 x 4 digit integer fields followed by 24 characters of text.

1.	Feature code (key)	I4
2.	Type of output (see below)	I4
3.	Pen/projector number	I4
4.	Variable, controlled by field 2 entry	I4
5.	Feature description	6A4

Integer entries are right-justified with leading zeroes or spaces, text entries must be space filled to 24 characters.

Type of output

Fields 2 and 4 are entered as follows·

Output type	Field 2	Field 4
Solid Line	1	Zero
Pecked Line	2	Length of peck (1:10 mm unit
Name	3	Zero
Centred symbol	5	Code in plotter software
Drawn symbol	6	Reference in symbol table
Flashed symbol	7	Zero

Header/terminator

The table is headed by the two digit code 03, and terminated by a five field feature record with feature code -9 and other fields zero or space filled.

TEXT TABLE (Option code 05)

For any height of text input for a name feature, a modified height may be substituted for plotting, and/or a specific pen or projector number selected.

The table is optional, but if included in the parameters, all name features will use it. If an exact equivalent to input size is not available as a key, the entry for the next largersize will be used from the table.

Table format

The table is headed by the two digit option code (05) and terminated by a three integer entry, as for a height, with the values -9, 0, 0.
A maximum of 20 height entries is permitted.

Each height entry consists of three fixed length integer fields (3I4)

Field 1 Key - an input height in units of .1mm

2 Modified (or unchanged) height, units as field 1

3 Pen/projector number

Height range

Heights on Ordnance Survey data will fall in the range 10-58 units

Example

Three entries in a text table require reduced text heights, the largest is to be plotted with a different pen.

```
05

001900150001
002400170001
002900210002

  -900000000
```

2.3.1 (Contd) Example

A feature description table consists of three features as follows:

a. Feature code 4, type 2, pen 3, length of peck 1.8 mm.

b. Feature code 28, type 3, no pen entry.

c. Feature code 37, type 5, pen 1, plotter symbol code 15.

03

Field 1	2	3	4	5					
4	2	3	18	BUIL	DING	(OPE	N SI	DED)	
28	3	0	0	NAME	/NUM	BER			
37	5	1	15	TELE	PHON	E CA	LL B	OX (G	PO)
-9	0	0	0						

Internal format of table

The parameters for each code are stored in a location whose start address is derived by program from the feature code, the code itself is not retained.

The first location for each code is used for a rejection indicator, if set by an option 10 entry.

2.3.3 SYMBOL TABLE (Option code 6)

The symbol table will hold the data necessary to construct symbols designed by the user.

Each symbol entry is prefaced by a numeric key by which the symbol may be called from the Feature Description table, or from other symbols in the table. A series of function codes and parameters are given overleaf, defining the form of the symbol, followed by the terminator 8888.

Each entry (key, function, or parameter) is written on a separate line.

The table is terminated by the code 9999.

A maximum of 20 symbols is permitted, overall table size must not exceed 200 entries.

Example: A symbol table of one symbol is to be set up, the symbol (key 14) consists of a square of side 5mm drawn around a central origin in a clockwise direction.

```
  06
  14
   4
 -25
 -25
   7
  25
   6
  25
   7
 -25
   6
 -25
   1
8888
9999
```

The pen position at start of symbol is assumed to be the origin, all subsequent positions to which moves are made are relative to this origin, given as X or Y values in units of 1/10 mm.

2.3.3 (Contd)

FUNCTION CODES USED IN DRAWN SYMBOL TABLES

DESCRIPTION	FUNCTION CODE	PARAMETER 1	PARAMETER 2
End of symbol	8888	-	-
Invisible move to symbol origin	1	-	-
Invisible move in x	2	Distance (1/10mm)	-
Invisible move in y	3	Distance	-
Invisible move in x/y	4	Distance x	Distance y
Visible move to origin	5	-	- -
Visible move in x	6	Distance	-
Visible move in y	7	Distance	-
Visible move in x/y	8	Distance x	Distance y
Pen/projector number	9	Number	-
Flash symbol	10	-	-
Scale factor	11	Factor X 1000	-
Compute scale factor	12	drawing distance (1/10 mm)	-
Jump to symbol n in table	13	n	-
Return to calling symbol	14	-	-

Notes:

Code 12 Scale factor is computed as $\frac{\text{Distance input from data tape}}{\text{Drawing distance}}$

To draw a symbol already constructed in the table at a different size, enter scale factor (code 11) followed by code 13.

The Bench Mark symbol, symbol 26 in the latest version of OSTESTPARAMS is oriented thus "<". On some early maps sheets this may lead to the Bench Mark being transposed by 180°. This can be corrected by altering the symbol tables entry to orient the symbol thus ">".

2.3.4 FEATURE LISTING (Option code 7)

1. Each page is headed by a title, sheet reference and plot scale, and column headings.

2. Features selected for each sheet used are listed as they occur on the input tape showing serial number, start and end NG co-ordinates and text information (if any).

3. Feature code and description is printed at each change of code.

4. End of sheet and end of run messages output as appropriate.

5. No error messages appear on the report.

6. The print layout requires a maximum of 80 characters per line, and 60 lines per page.

A sample page is shown overleaf.

·3·4 (contd)

IGITAL MAPPING - D09 PLOT LISTING

HEET REF 333000. 226000. PLOT SCALE 2500.

EATURE CODE	FSN	START CO-ORDS		(TEXT OF NAME)	END CO-ORDS	NAMF CAT	TEXT HEIGHT
28 NAME							
	1	333956.0	226235.2	P]ARSONAGE		1	1.7
	2	333955.9	226229.1	F]ARM		1	1.7
	3	333939.6	226071.8	I]SSUES		9	1.7
	4	333980.8	226057.8	I]SSUES		9	1.7
	5	333776.8	226254.5	I]SSUES		9	1.7
	6	333534.9	226233.3	S]PRING		9	1.7
	7	333535.3	226225.8	P]OND		7	1.7
	8	333509.4	226223.2	BM		7	1.7
	9	333462.3	226223.4	M]IDDLE		1	1.9
	10	333462.5	226217.1	H]UNTHOUSE		1	1.9
	11	333462.4	226209.9	F]ARM		1	1.9
	12	333421.1	226176.7	I]SSUES		9	1.7
	13	333496.2	226123.3	W]ELSH$ H]UNTHOUSE		1	1.9
	14	333573.0	226087.1	P]ATH		9	1.7
	15	333520.9	226027.3	I]SSUES		9	1.7

Note: In text,] indicates lower case to follow, $ upper case to follow

2.3.4 (contd)

DIGITAL MAPPING - D09 PLOT LISTING

SHEET REF 333000. 226000. PLOT SCALE 2500.

FEATURE CODE	FSN (TEXT OF NAME)	START CO-ORDS		END CO-ORDS		NAME CAT	TEXT HEIGHT
30	FENCE,WALL (NON ROAD) .						
	283	333492.5	226135.5	333487.8	226138.6		
32	SURV PECKS (NON ROAD) .						
	284	333484.2	226112.4	333698.0	226046.9		
	285	333483.6	226111.1	333699.3	226045.1		
64	SINGLE STREAM .						
	286	333563.7	226670.6	333267.5	226425.6		
58	POINT FEATURE (CIRCLE) .						
	287	333564.3	226671.1				
30	FENCE,WALL (NON ROAD) .						
	288	333098.2	226568.6	333000.3	226545.2		

END OF SHEET

END OF RUN

2.3.5 REJECT BY NAME CATEGORY (option code 11)

Each text feature input has one of the following category codes

- 0 Street name
- 1 Building name
- 2 Object shown by dot or circle
- 3 Area name, town, village etc
- 4 Archaeological names
- 5 Boundary information
- 6 House numbers
- 7 Ground survey/topo names
- 8 Road numbers
- 9 Miscellaneous

A table set up by the program will hold a zero entry for each category.

If option 11 is used, each category for rejection is set to 1, features with these category numbers will then be omitted from the plot.

The rejections set remain in force for all plots until cancelled by another code 11 entry.

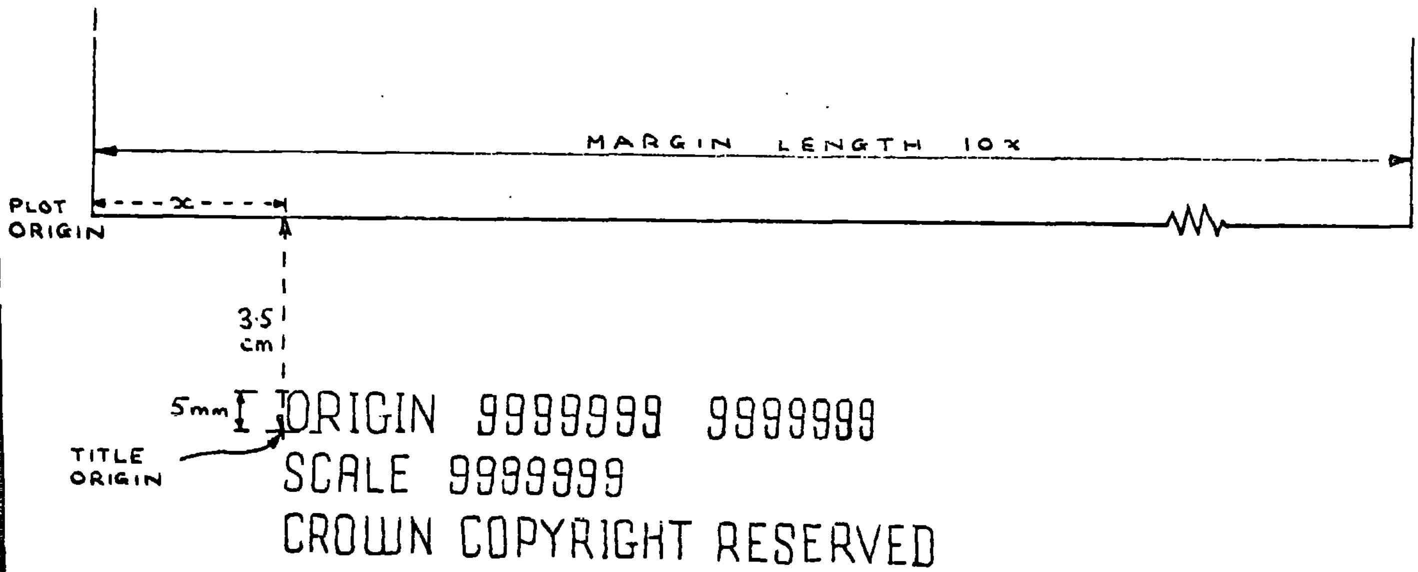
2.3.6 TITLE OF PLOT (option code 15)

The title block, including NG co-ordinates of the SW corner of the plot, plot scale and a copyright note, is laid out as shown below. The title is output as a basic part of the plot program.

If code 15 is not used, the title origin will be set as shown below in relation to the SW corner of the plot, with a text height of 5mm.

The use of code 15 allows a choice of origin, relative to plot origin, and text height.

If codes 18 or 19 (plot origin location and orientation) are used, the title location will be relative to the SW corner of the plot, the title will be rotated if necessary to remain parallel with the south (north) margin.



2.3.7 MARGINAL DATA (option code 16)

Up to 80 characters of specified text may be plotted at any angle and location in or around the plot area, text height and pen/projector number are also specified.

The controlling parameters are read as six fixed length fields in one record as follows (numbers in brackets give field length)

- Pen/projector number (4)
- Start of text location relative to plot origin (X,Y ±mms) (2x8)
- Height of text (1/10mm units) (6)
- Angle of text, anticlockwise from horizontal, 1/10 degrees (6)
- Number of characters to follow (4)

The corresponding text follows in four character records, maximum 20 records, the last is space filled up to four characters.

Example:

The name CITY OF LANCASTER is required located centrally, and parallel to the east margin of a plot of 600mm x 600mm. Required text height is 8mm and pen number is 2.

```
16
0002-00000100000025000008000009000020
CITY
 OF
LANC
ASTE
R
```

Each entry must be preceded by code 16.

3 OUTPUT DATA - INTERFACE WITH USER ROUTINES

A common area in the plot program, named INFACE, consists of 33 fortran units of storage. Up to 14 items of data may be placed in fixed positions in this area (layout overleaf).

Different combinations of items will be present according to output type, this is signified by a code in the first position in INFACE.

Each group of items provides data from which parameters for a call to a plotter routine can be generated.

No data is present with output codes 6 or 8, these are merely triggers to the user's start and end of plot routines.

Data in the area INFACE will be accessed through a routine named PLOTTR - to be written by the user.

Output codes

1 Move to given point (pen up/down)

2 Plot number

3 Plot text

4 —

5 Draw centred symbol

6 End of plot

7 Flash symbol

8 Start of plot

Location	Description	Format (integer/ real)	Value range	Used for output types
1	Output code	I	1-8	All
2	Pen/projector number	I	Set in Feature Description table*	All except 6,8
3	Draw/non draw indicator	I	0 = pen up/proj off, 1 = pen down/proj on	1,5
4	Symbol code	I	Set in Feature Description table	5
5	Number of text characters	I	1-80	3
6	Text characters	A	Letters, numbers, punctuation	3
26 27	)Co-ordinates)	(R (R	X) relative to plot origin (mms) Y)	All except 6,8
28	Numeric value for plotting	R		2
29	Text character/symbol height	R	In millimetres	2,3,5
30	Angle of text/symbol	R	In degrees, anti-clockwise from horizontal	2,3,5
31	Text style	I	1-9	)Of no value to
32	Name category	I	1-10	)user
33	Distance	R	In millimetres	)

* Not entered for pen-up moves.

.1 SUBROUTINE PLOTTR

A version of this subroutine, for CALCOMP plotters, is appended to the source listing on printer and paper/magnetic tape.

It should provide the link with the data output to the common area, and includes calls to CALCOMP plotter software.

In its basic form it is suitable for users with an ICL installation and a Calcomp imperial plotter.

Amendments for other users are as follows

IBM

Label 300 : Parameter LTEXT is an array, the subscript may need to be removed.

Label 800 : As above for IBUF

Metric Calcomp

Statement after label 800

For plotters working in centimetres, change parameter to (0.1)

DIMENSION statement

This statement is unnecessary for certain plotters.

4 ERROR ACTION

Errors associated with the logic of the program will cause a halt (non-restartable) with one of the following values output to the console or operating system monitor listing. The value is preceded by the word STOP.

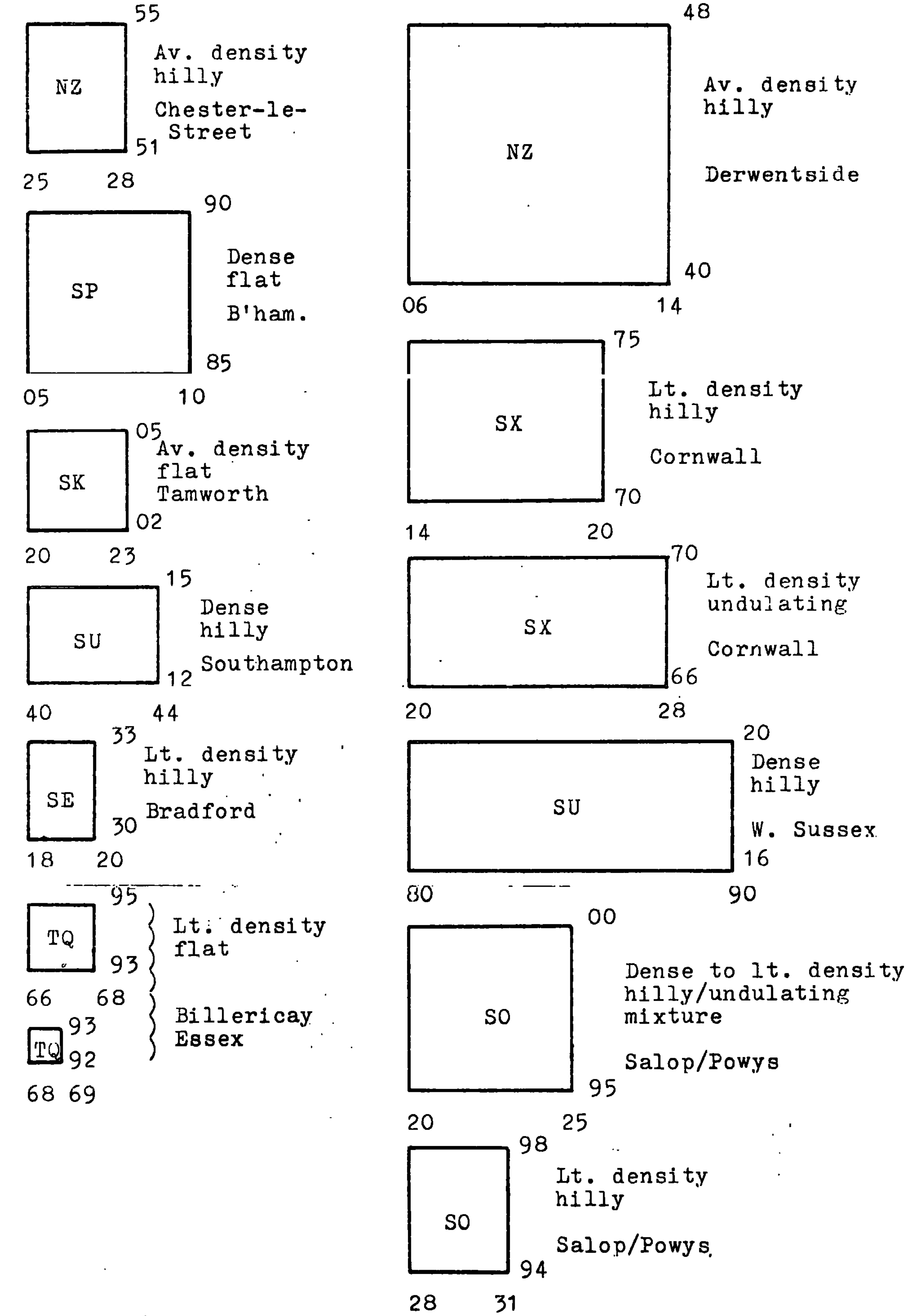
100	Parameter type not in range 01-24
201	Minimum essential parameters (01 or 02 and 03) not input
102)	Plot serial number option not permitted in a part sheet plot (02)
170)	
103	Feature description table has nil, or more than 150 entries
105	Text table has more than 20 entries
106	More than 20 symbol types in symbol table
107	More than 200 entries in symbol table
1113	Symbol code set in FDT not in symbol table
110	More than 150 feature codes rejected (code 10)
111	More than 10 name categories rejected (code 11)
120	More than 30 serial numbers rejected (code 12)
130	More than 30 serial numbers selected (code 13)

Errors associated with input and output from/to peripherals will cause the Fortran error action laid down at the users' installation.

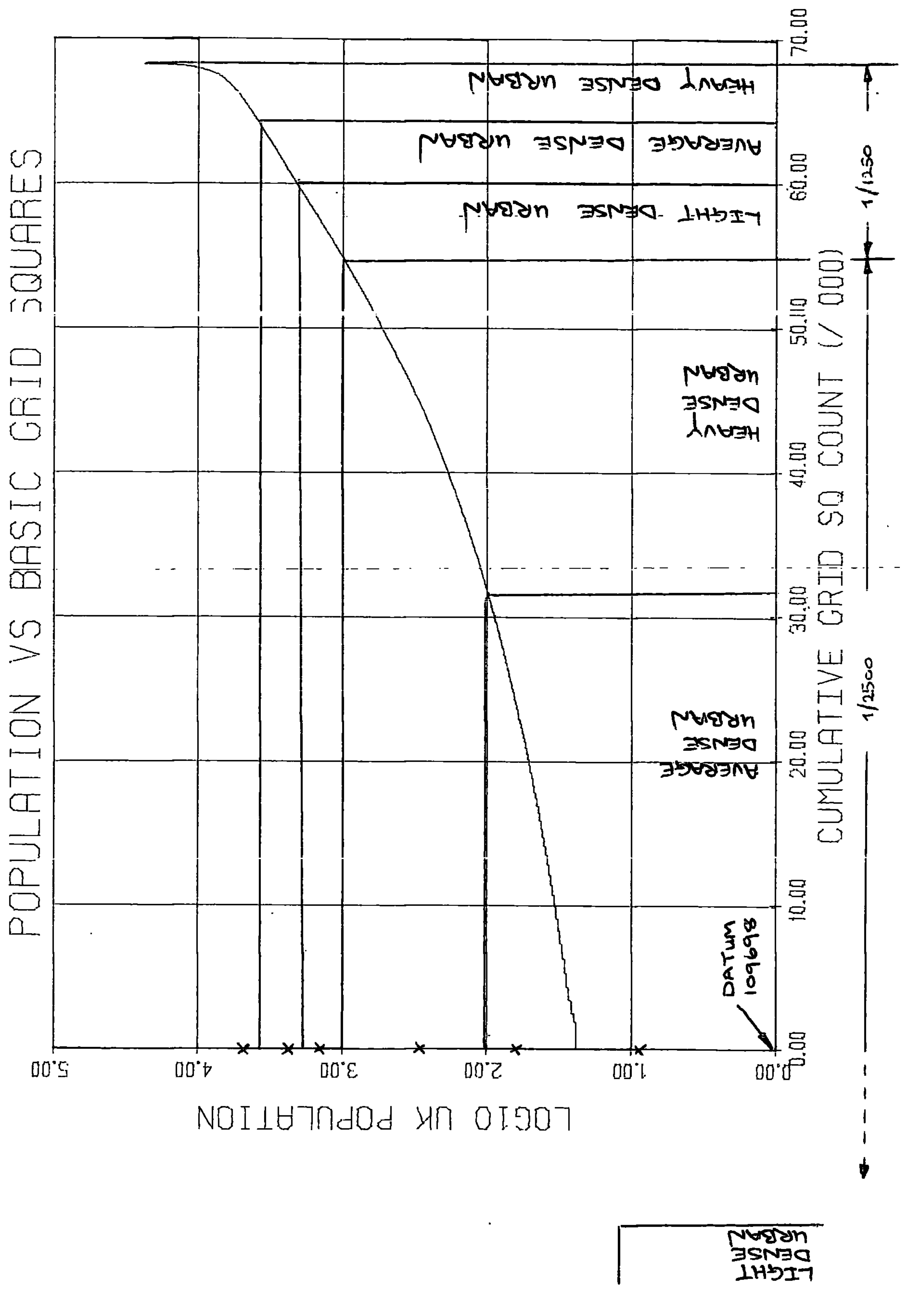
Appendix E

1. 1:1250 scale

2. 1:2500 scale



<u>Appendix F</u>



Appendix G

Projected Figures For The UK 1:1250 Digital Database

Part a - DMC Sequence Code Frequencies

-1	Start of sheet	53,316	
-2	End of sheet	53,316	
-3	End of file	N/A	
-4	Start of feature	68,587,521	(19,506)
-5	End of feature	68,587,521	(19,506)
-6	Not used		
-7	Features (Line)	56,182,495	(16,224)
-8	Features (Text, Symbol)	12,405,025	(3,755)
-9	Grid square indicator	111,946,815	(28,796)
-10	Text classification	10,610,854	(3,210)
-11	Character calls	10,610,854	(3,210)
-12	No. of orientations	68,599,511	(19,501)
-13	Invisible line flag	2,200,606	(1,496)
-14	Not used		
-15	Distance calls	7,023	(13)
-16	Not used		
-17	Not used		

(figures in brackets are the ± margins of error at 95% CL)

Part b - Feature Frequency and breakdown

F/C	FREQ	No. PTS	No. LINE SEGMENTS	PLOT TOT LINE LENGTH m	GROUND EQUIV km	SEGMENTS MEAN	SEGMENTS MAX	SEGMENTS MIN
1	173783 155	2590244 2487	2416461 2357	11666.1 9.9	14582.7 12.4	5.7	122	0.0
2	4891671 2584	21502146 13465	16610476 11773	43776.4 21.2	54720.5 26.5	3.1	127	0.0
3	9167217 3261	78899110 31959	69732398 29235	343223.0 108.6	429028.8 135.8	5.8	424	0.0
4	2145621 1110	5930148 3109	3784527 2123	13431.1 8.2	16788.9 10.2	4.6	160	0.0
5	24082 58	53570 125	29488 70	179.6 0.5	224.5 0.6	8.7	59	.14
6	329967 288	667811 585	337845 298	2600.6 2.4	3250.8 3.0	9.7	114	.59
7	4149 12	176136 446	171987 436	726.0 1.9	907.5 2.4	5.3	116	.06
8	13337 32	678139 1060	664802 1033	2810.6 3.6	3513.3 4.5	5.1	106	0.0
10	9337 15	404554 848	395216 840	2528.7 3.8	3160.9 4.8	8.9	164	0.0
11	207 1	207 1						
12	30606 40	30606 40						
13	15123 24	15123 24						
14	2315 16	42020 223	39706 218	178.1 1.1	222.6 1.4	5.9	127	0.0
15	428385 680	5157911 9965	4729526 9328	45781.0 74.3	57226.3 92.9	11.2	512	0.0
16	2382 17	15302 130	12920 120	83.6 0.7	104.5 0.9	8.8	30	.09
18	415 3	10373 74	9958 71	140.1 1.0	175.1 1.2	17.6	87	.14
19	2663 13	38657 75	32993 64	625.4 1.4	781.8 1.7	28.9	209	.63
20	116917 316	235588 634	118671 318	136.1 0.3	170.1 0.4	1.4	5	0.0
21	1684816 539	54388673 25360	52703857 25087	182371.5 46.5	227964.4 58.1	4.2	325	0.0
23	40051 32	896203 1137	856152 1114	4163.2 3.4	5204.0 4.3	5.2	201	0.0
24	313648 259	313648 259						
25	4412 6	4412 6						

...,cont.

F/C	FREQ	No. PTS	No. LINE SEGMENTS	PLOT TOT LINE LENGTH m	GROUND EQUIV km	SEGMENTS MEAN	SEGMENTS MAX	SEGMENTS MIN
26	198885	198885						
	101	101						
27	449647	449647						
	197	197						
28	10613986	10610506						
	3211	3210						
29	2891655	26839424	23947770	119424.2	149280.3	6.1	215	0.0
	2058	16286	14673	55.8	69.8			
30	26257922	92696002	66438079	435777.6	544722.0	8.0	238	0.0
	9623	28227	19638	134.8	168.4			
31	1387068	17509898	16122830	52476.4	65595.5	4.1	204	0.0
	602	9834	9411	24.5	30.6			
32	3440823	33191738	29750915	112126.0	140157.5	4.4	403	0.0
	1308	13781	12717	41.7	52.1			
33	30593	145468	114876	1151.2	1439.0	11.7	504	0.0
	85	444	378	3.9	4.8			
34	54348	157501	103152	808.6	1010.8	9.9	55	0.0
	96	298	216	1.5	1.9			
35	71054	1951212	1598357	4652.8	5816.0	2.8	113	0.0
	66	2584	2529	4.9	6.2			
36	7965	265247	257282	605.4	756.8	2.2	36	0.0
	13	636	625	1.0	1.2			
37	53449	53449						
	44	44						
44	141	562	422	26.4	33.0	78.3	109	38.
	1	6	4	0.3	0.3			
45	422	8852	8430	10.4	13.0	2.0	15	0.2
	4	88	86	0.1	0.1			
47	11689	42848	31159	1889.2	2361.5	91.4	325	1.1
	19	69	53	3.2	4.1			
48	505	505						
	1	1						
49	7023	7023						
	13	13						
51	7469	7469						
	38	38						
52	1374295	5786101	4411805	5169.0	6461.3	1.5	39	0.0
	935	4107	3351	4.0	5.0			
54	556	556						
	3	3						
55	622	16390	15767	164.5	205.6	20.9	506	0.5
	4	109	108	1.2	1.5			
57	617610	617610						
	616	616						
58	4935	4935						
	17	17						

...cont.

F/C	FREQ	No. PTS	No. LINE SEGMENTS	PLOT TOT LINE LENGTH m	GROUND EQUIV km	SEGMENTS MEAN	SEGMENTS MAX	SEGMENTS MIN
59	85519	6061688	5976169	8366.4	10458.0	1.9	174	0.0
	90	6930	6866	8.5	10.6			
60	34573	846008	811435	3318.1	4147.6	5.3	115	0.0
	74	1814	1762	6.3	7.8			
61	25449	1321496	1296047	3109.3	3886.6	2.4	57	0.0
	45	3511	3477	6.4	8.0			
62	23527	1056017	1036986	1317.3	1646.6	1.3	50	0.0
	24	2765	2751	2.3	2.8			
64	52037	1711930	1667059	3059.4	3824.3	2.0	76	0.0
	47	1550	1517	2.4	2.9			
65	44807	1166498	1121691	2960.3	3700.4	3.5	132	0.0
	63	1757	1703	4.6	5.8			
66	13776	749042	735267	1696.5	2120.6	5.2	152	0.0
	33	1024	1007	3.0	3.8			
67	415	415						
	2	2						
68	9957	9957						
	17	17						
69	33940	33940						
	27	27						
70	45809	45809						
	56	56						
71	9695	479261	469566	1215.8	1519.8	3.2	73	0.0
	45	1812	1783	4.2	5.2			
72	1405	299696	298291	374.0	467.5	1.6	19	0.0
	11	2234	2223	2.8	3.5			
78	22457	861576	839119	5512.2	6890.3	8.0	185	0.0
	19	1155	1144	4.9	6.2			
79	281	5480	5199	54.9	68.6	13.2	90	.14
	2	55	53	0.5	0.7			
80	9557	235381	225824	943.9	1179.9	5.8	90	.06
	17	374	359	1.7	2.1			
81	20719	614381	593661	1965.6	2457.0	4.2	90	0.0
	28	998	972	3.3	4.1			
83	539361	4831101	4291740	19055.5	23819.4	5.4	115	0.0
	373	3069	2730	12.3	15.4			
84	3817	165820	162003	203.4	254.3	1.3	25	0.0
	9	674	666	0.6	0.8			
85	1265	12224	10959	174.3	217.9	19.9	189	.13
	8	74	67	1.0	1.3			
86	843	5901	5058	108.4	135.5	26.8	211	1.1
	5	47	44	0.7	0.9			
88	703	9554	8852	59.7	74.6	8.4	58	1.1
	5	72	67	0.5	0.6			

...cont.

F/C	FREQ	No. PTS	No. LINE SEGMENTS	TOT LINE LENGTH m	GROUND EQUIV km	SEGMENTS MEAN	SEGMENTS MAX	SEGMENTS MIN
89	11688	294209	2822521	2110.4	2638.0	9.2	141	0.0
	27	791	768	4.5	5.6			
90	28855	850737	826682	6510.3	8137.9	9.5	183	0.0
	35	1081	1051	6.3	7.9			
91	1489	96774	95285	196.8	246.0	2.3	43	0.0
	6	449	444	0.9	1.1			
92	1753	40774	39021	225.8	282.3	6.9	61	.14
	7	193	186	0.9	1.1			
93	17523	447375	429852	3439.5	4299.4	11.4	307	0.0
	27	615	593	4.8	6.0			
94	1201	81952	80751	137.8	172.3	2.3	35	0.0
	4	303	299	0.5	0.7			
95	984	4918	3934	67.1	83.9	21.3	39	7.5
	5	26	22	0.4	0.5			
96	49136	1142707	1093571	9462.5	11828.1	11.1	253	0.0
	71	1545	1484	10.3	12.9			
97	64628	477915	462422	2697.6	3372.0	7.0	173	.06
	39	1212	1180	4.5	5.6			
98	487219	6874437	6387218	48339.6	60424.5	10.9	287	0.0
	231	3599	3417	24.4	30.4			
99	1050	77667	76616	66.8	83.5	1.2	6	.06
	5	327	323	0.3	0.4			
100	18395	393063	374668	2763.5	3454.4	11.2	318	0.0
	37	817	793	5.5	6.9			
101	5138	53301	48163	924.4	1155.5	19.8	409	.28
	20	328	309	5.3	6.7			
102	17028	317555	300528	2147.0	4294.0	8.2	190	0.0
	85	1822	1744	10.9	13.6			
103	13700	27398	13700	182.9	228.6	20.5	68	2.1
	72	145	72	1.5	1.8			
104	141	281	141	0.2	0.2	1.3	1.3	1.3
	1	3	1	0.0	0.0			
108	49634	676343	749735	1779.0	2223.8	3.0	64	0.0
	160	2561	2431	5.1	6.5			
109	422	843	422	0.5	0.6	1.5	2	1.3
	5	9	5	0.0	0.0			

(smaller figures below actual projections are margins of error at the 95% CL)

Part c - other information

Total no. of text characters	59,438,113	(18,189)
Total (-15 coded) distance	454km (gnd)	(1.2km)
Total inked in line segments	324,349,181	(79,324)
Total invisible segments	2,200,606	(1,496)
Total segments generated	326,549,786	(80,820)

Total distance
generated by lines: 1,903,547km ground (421.2km)
1,522,838m map

No. of DMC records	940,757,829	(235,982)	(251 tapes)
No. of points	395,133,827	(95,906)	
No. of ignored codes	413,710	(279)	
No. of ignored records	12,140,911	(13,848)	(DMC - 8 bytes)

Section 2

Projected Figures For The 1:2500 Digital Database

Part a - DMC Sequence Code Frequencies

-1	Start of sheet	164,461	
-2	End of sheet	164,461	
-3	End of file	N/A	
-4	Start of feature	47,782,037	(34,050)
-5	End of feature	47,782,037	(34,050)
-6	Not used	-	
-7	Features (Line)	35,663,222	(27,834)
-8	Features (Text, Symbol)	12,118,817	(6,630)
-9	Grid square indicator	98,029,281	(48,225)
-10	Text classification	7,953,958	(5,681)
-11	Character calls	7,953,958	(5,681)
-12	No. of orientations	47,913,728	(34,050)
-13	Invisible line flag	173,153	(542)
-14	Not used		
-15	Distance calls	no data to project	
-16	Not used		
-17	Not used		

(figures in brackets are the ± margins of error at 95% CL)

Part b - Feature Frequency and breakdown

F/C	FREQ	No. PTS	No. LINE SEGMENTS	PLOT TOT LINE LENGTH m	GROUND EQUIV km	SEGMENTS MEAN	SEGMENTS MAX	SEGMENTS MIN
1	36061	374228	338167	727.4	1818.5	5.6	37	.13
	102	1325	1230	3.0	7.5			
2	2163873	10058476	7894604	13898.9	34747.3	4.4	43	.06
	3369	15189	11835	19.6	49.1			
3	3494897	24237304	20742406	58720.5	146801.3	7.1	82	.06
	4742	35680	30971	81.1	202.7			
4	701211	1876886	1175675	3834.1	9585.3	8.1	50	.06
	581	1556	991	3.1	7.6			
6	5528	11056	5528	14.7	36.8	6.1	8	3.4
	53	107	53	0.1	0.3			
7	128489	13526165	13397675	31513.9	78784.8	5.4	285	.06
	144	18659	18568	30.5	76.2			
8	21768	11665358	1643590	4858.7	12146.8	7.7	119	.06
	54	5184	5146	10.9	27.3			
9	9729	1583701	1573973	3149.0	7872.5	4.3	91	.06
	41	8854	8820	11.9	29.8			
11	1337	1337						
	10	10						
12	274507	274507						
	322	322						
13	150923	150923						
	232	232						
15	3568	104529	100961	886.1	2215.3	21.9	519	0.2
	39	1276	1240	9.5	23.9			
19	18876	524967	506092	2819.0	7047.5	16.9	87	.13
	61	1753	1699	10.9	27.3			
21	2021219	46049867	44028648	133618.6	334046.5	7.5	292	.14
	1261	29545	28506	75.2	187.9			
22	3324	55437	52113	151.6	379.0	8.7	35	.14
	24	553	533	2.0	5.0			
23	129226	4626074	4496848	10230.6	25576.5	5.9	211	.06
	363	13727	13408	28.5	71.3			
25	6111	6111						
	25	25						

F/C	FREQ	No. PTS	No. LINE SEGMENTS	PLOT TOT LINE LENGTH m	GROUND EQUIV km	MEAN	MAX	MIN
26	483108 248	483108 248						
27	1170146 585	1170146 585						
28	7963593 5677	7953601 5678						
29	2541061 2362	86840958 57901	84299897 56085	204225.4 98.0	510563.5 245.1	6.3	472	.06
30	15732910 13108	291401829 154096	275668920 144628	691527.9 199.9	1728819.8 499.9	6.4	394	.06
31	941967 ~~1102~~	18247824 ~~21274~~	17305857 ~~20307~~	39052.7 ~~40.7~~	97631.8 ~~101.7~~	5.4	318	.06
32	1180469 1130	19464373 18143	18516151 17355	45582.5 51.2	113956.3 128.1	6.4	263	.06
33	5708 117	11416 249	5708 133	11.0 0.5	27.56 1.2	8.7	31	.61
34	714 15	1427 30	714 15	1.4 0.03	3.6 0.07	5.0	5	4.9
35	1109929 550	36157092 22140	35767162 21700	60802.8 34.2	152007.0 85.6	4.3	279	.06
36	458915 1354	15090910 32743	14631995 31653	14482.8 36.8	36207.0 91.9	2.5	58	.06
37	472013 36	472013 36						
39	980 7	980 7						
44	2163 18	13794 84	11631 71	47.9 0.3	119.8 0.8	13.9	30	3.9
45	1183 23	5917 58	4734 37	42.5 0.3	106.21 0.8	21.3	44	5.6
47	2497 59	8562 141	6065 88	254.9 6.9	637.3 17.4	156.7	378	.54
48	714 10	714 10						
51	5708 94	5708 94						
52	160368 673	563506 1810	403138 1213	245.0 0.8	612.5 1.9	1.5	5	.31

F/C	FREQ	No. PTS	No. LINE SEGMENTS	PLOT TOT LINE LENGTH m	GROUND EQUIV km	SEGMENTS MEAN	SEGMENTS MAX	SEGMENTS MIN
54	8581	8581						
	38	38						
55	83347	2369975	2286627	2279.2	5698.0	2.8	287	.06
	530	12004	11487	8.7	21.7			
57	197477	197477						
	377	377						
58	73882	73882						
	88	88						
59	648828	56916437	56267609	61237.7	153094.3	2.7	186	.06
	555	52758	52388	53.7	134.4			
61	17109	621686	604576	1151.9	2879.8	4.8	48	.06
	71	2375	2326	4.7	11.8			
62	219511	15128139	12908629	9031.2	22578.0	1.7	53	.06
	251	18531	18362	14.0	35.0			
64	1571041	73025199	71454159	82186.5	205466.3	3.0	257	.06
	788	37511	37018	37.1	92.7			
65	124989	2476317	3101688	5677.7	14194.3	5.7	131	.06
	207	5839	5668	11.7	29.2			
66	97300	9358593	56261292	11911.2	29778.0	3.2	140	.06
	176	17297	17162	21.2	53.1			
67	357	357						
	7	7						
68	5818	5818						
	19	19						
69	878967	878967						
	407	407						
70	672521	672521						
	545	545						
78	980	1960	980	0.4	1.0	1.0	1	1.0
	7	15	7	.0	.0			
80	363498	12263465	11899969	31633.6	79084.0	6.9	496	.06
	591	23378	22847	66.2	165.4			
81	813498	26182150	25368652	60855.1	152137.8	6.1	218	.06
	1032	50279	49393	123.2	308.1			
82	216603	216603						
	190	190						
83	29697	281792	252095	437.8	1094.5	6.0	42	.06
	207	2473	2274	6.3	15.6			

F/C	FREQ	No. PTS	No. LINE SEGMENTS	PLOT TOT LINE LENGTH m	GROUND EQUIV km	SEGMENTS MEAN	SEGMENTS MAX	SEGMENTS MIN
84	100541	4832875	4732333	4722.8	11807.0	2.2	65	.06
	384	18593	18244	15.3	38.4			
90	9269	355007	345738	3087.4	7718.5	20.5	162	.13
	25	1237	1216	9.2	22.9			
93	12849	513243	500395	3692.9	9232.3	18.7	144	.13
	53	2145	2096	13.6	34.1			
96	98350	5620861	5522513	22846.8	57117.0	10.4	233	.06
	127	6989	6877	25.2	62.9			
97	141787	8982748	8840962	27147.5	67868.8	8.1	221	.06
	148	9886	9761	28.7	71.7			
98	419303	14114953	13695650	37076.9	92692.3	7.0	211	.06
	512	14852	14470	35.9	89.9			
101	1784	49946	48162	443.0	1107.5	23.0	520	.14
	20	594	575	4.8	11.9			
103	16601	33201	16601	133.2	333.0	19.2	42	6.8
	63	125	63	0.5	1.3			
108	11671	245918	234246	404.7	1011.8	4.8	33	.13
	46	939	901	1.5	3.7			

(smaller figures below actual projections are margins of error at the 95% CL)

Part c - other information

Total no. of text characters	65,659,809	(41,189)
Total (-15 coded) distance	no data to project	
Total inked in line segments	768,765,365	(335,963)
Total invisible segments	173,153	(542)
Total segments generated	768,938,518	(336,504)

Total distance
generated by lines: 4,215,581km ground (1,219km)
1,686,232m map

No. of DMC records	1,239,969,874	(587,919)	(331 tapes)
No. of points	816,710,565	(362,338)	
No. of ignored codes	370,035	(381)	
No. of ignored records	25,294,863	(27,849)	(DMC - 8 bytes)