THE APPLICATION OF REMOTE SENSING TO WATER RESOURCES

ROBERT SUTHERLAND DRAYTON

Doctor of Philosophy

THE UNIVERSITY OF ASTON IN BIRMINGHAM

February 1989

The University of Aston in Birmingham

The Application of Remote Sensing to Water Resources

Robert Sutherland Drayton

Doctor of Philosophy 1989

Synopsis

Techniques are developed for the visual interpretation of drainage features from satellite imagery. The process of interpretation is formalised by the introduction of objective criteria. Problems of assessing the accuracy of maps are recognized, and a method is developed for quantifying the correctness of an interpretation, in which the more important features are given an appropriate weight. A study was made of imagery from a variety of landscapes in Britain and overseas, from which maps of drainage networks were drawn. The accuracy of the mapping was assessed in absolute terms, and also in relation to the geomorphic parameters used in hydrologic models. Results are presented relating the accuracy of interpretation to image quality, subjectivity and the effects of topography. It is concluded that the visual interpretation of satellite imagery gives maps of sufficient accuracy for the preliminary assessment of water resources, and for the estimation of geomorphic parameters.

An examination is made of the use of remotely sensed data in hydrologic models. It is proposed that the spectral properties of a scene are holistic, and are therefore more efficient than conventional catchment characteristics. Key hydrologic parameters were identified, and were estimated from streamflow records. The correlation between hydrologic variables and spectral characteristics was examined, and regression models for streamflow were developed, based solely on spectral data. Regression models were also developed using conventional catchment characteristics, whose values were estimated using satellite imagery. It was concluded that models based primarily on variables derived from remotely sensed data give results which are as good as, or better than, models using conventional map data. The holistic properties of remotely sensed data are realised only in undeveloped areas. In developed areas an assessment of current land-use is a more useful indication of hydrologic response.

KEYWORDS: Remote sensing, hydrology, modelling

ACKNOWLEDGEMENTS

I would like to acknowledge the invaluable guidance given to me by my supervisor, Dr. T.R.E. Chidley, and the continued enthusiasm shown by my Head of Department, Professor H.R. Evans. My thanks also go to Mrs. M. Ellis, who as always, has produced work of the highest standard.

Above all else I must express my gratitude to my family, whose loving support and encouragement have made this work possible. Thank you Sheila, Mary, Eleanor, Sarah and Caroline.

CONTENTS

PAGE NO

PAGE NO.

LIST OF TABLES

LIST OF FIGURES

PART 1 - INTRODUCTION

PART 1 CHAPTER 1 GENERAL INTRODUCTION

The theme underlying this work is one of optimism, that the latest advances in space technology may be used beneficially in an activity which is fundamental to the needs of mankind - the management and development of natural water resources.

Given an adequate supply of reasonable quality, people need spend less time in fetching water, and will suffer less from water-related disease. Given more time and better health, then food productivity increases, which promotes still better physical well-being and better resistance to disease, with the result that people are set on an upward spiral leading to social and economic development.

Sadly, in many parts of the world, the basic water resources are not provided. This is due not to lack of effort or determination, but simply to a lack of information. Maps of the river systems may be available only at 1:1 million or 1:250,000 scale, and are frequently unreliable. The existence of lakes and reservoirs is often not known to the planners, while rivers may change their course with a rapidity unappreciated in our temperate climate. The hydrological response of catchments can be radically altered by the process of deforestation or slash-and-burn cultivation over large areas, which cannot be monitored using traditional methods of survey. In the absence of this kind of information it is impossible to devise any sensible strategy for the development of water resources.

Even though this situation still prevails, great advances have been made in acquiring information on the world's surface. Following the development of orbiting satellites in the 1960's and 1970's we now have a second generation of earth resources satellites, which have achieved an operational, commercial status. They are capable of observing all parts of the globe on a regular basis, and provide subtle interpretations of conditions at the earth's surface.

The solution to the problem of hydrologic information thus exists, if the satellite data can be interpreted to give useful end products. One approach, which has been found valuable in the United States of America, is to interpret satellite data in terms of conventional maps, i.e. cropping patterns, soil types and land-use, which provide the inputs to conventional models.

The proposition which is examined in this thesis is that there is a better way of using satellite data, which is based on the idea that it is holistic, bringing together information on many aspects of the area being observed - its vegetation, soil type, soil moisture, slope and so on. This is quite different from the information presented in maps and used in conventional models, which is separated into distinct categories. Thus, to use satellite data merely to acquire conventional map interpretations is to ignore much of the richness of the data. However, to use spectral data in its own right requires that new models should be derived to take advantage of it.

The work presented here is a first step in that direction, in which the value of satellite data is examined in relation to the needs of water resource development, and simple models are developed. Recommendations are made for a second stage of work in which the benefits of the distributed nature of the data may be realised.

Some of the work presented in this thesis has already been published, and a list of references is given in Appendix 4.

PART 1 CHAPTER 2 AIMS AND OBJECTIVES

The principal aim of this work was to investigate whether remotely sensed data could be of value in the assessment of surface water resources. Clearly, it was not practicable to investigate all aspects of remote sensing, nor all aspects of water resource assessment. Therefore the aim of the study was refined to concentrate on the use of data provided by orbiting satellites, particularly those of the Landsat family, for which an adequate archive of data existed.

With regard to water resource assessment, the decision was made to limit the investigation to the techniques which would be useful at an early stage of survey, where little or nothing would be known about the surface water resources. In other words, the techniques which would be of most value in the developing world. Unfortunately, the paucity of information makes it difficult to develop models, so the decision was made to base the study on an area for which good quality data were available, and to develop models and techniques which are sufficiently robust to be of general value.

The two techniques which were considered to be of most value at an early stage of water resource survey were (i) mapping of the surface water resources (i.e. rivers and lakes) and (ii) modelling of key streamflow indices.

Thus, the objectives of the project were to :

1. Acquire maps of the surface water resources of the study area from conventional sources, and assess their validity.

2. Develop simple visual techniques for the interpretation of satellite imagery to provide maps of the surface water resource,
3. Develop a method of assessing the accuracy of the interpretations,
4. Acquire hydrologic data for the study area and select suitable streamflow indices,
5. Acquire satellite imagery and select appropriate variables which may be evaluated from the remotely sensed data,
6. Examine the correlation between hydrologic variables and remotely sensed variables, and hence propose simple models for estimating streamflow,
7. Test the models, and select simple robustmodels which would be suitable for general use.

In view of the clear split between the mapping work and the modelling work the main body of this report has been split into two parts dealing with each topic separately.

PART 1 CHAPTER 3 INTRODUCTION TO REMOTE SENSING

Remote sensing is the activity of observing phenomena at a distance, using emitted or reflected electromagnetic radiation. In the more limited definition used in this thesis, the activity of remote sensing is concerned primarily with observation of the earth's surface. The origins of remote sensing are in the military use of aerial photography but it has now been extended to include radar, infra-red photography, television and multi-spectral scanning from a wide range of platforms including satellites.

There are five major elements in the process : sensing, carrying the sensor, transmitting the data, processing and interpreting the data. The most common sensor is still the photogrpahic camera, carried in an aircraft, providing pictorial information based on visible and infrared radiation, which is interpreted visually. The exciting new developments, though, have come from the use of multi-spectral scanners carried in satellites. These sense the scene in a number of spectral bands simultaneously, and transmit the information in digital form. Land-based stations receive the data, which can then be analysed mathematically, or processed to create a wide range of pictorial interpretations. Recent advances in space technology and computing technology have made this an extremely powerful method for investigating problems, and the range of applications is limited only by our imagination.

In multi-spectral imaging a number of sensors are used, each sensor responding to a different spectral band (or channel). The scene is scanned simultaneously by all detectors, and the picture is built up from the series of stripes created by the forward motion of the carrier (usually an orbiting satellite).

Processing data in a number of bands is far more complex than in one band, but gives a richer body of information to use in interpretation. A multi-spectral scanner does not produce a photograph in a conventional sense, but produces numeric information on the degree of reflection in each spectral band, which is transmitted in digital form (after some on-board processing) to receiving stations on the ground. Analysis may be performed directly on the digital information, but it is common to create false colour images to examine visually. These may emphasise the reflection in individual bands, or the contrast between bands, or areas with a similar spectral "signature" i.e. similar intensity of reflection in all bands.

For example, a colour photograph may show two fields of the same crop as being green. In a multi-spectral image of the same area, the use of green combined with infrared would highlight areas of diseased or water-stressed crops, which would have high reflectance in green, but low reflectance in infrared. Healthy crops have a high reflectance in both bands. Thus by enumerating the contrast between green and infrared, the pattern of disease can be built up. The bands used for comparison would depend on the type of study being carried out, but once the scene has been scanned, the information on all bands is saved in the data archive and is available for use at any later time.

3.1 SENSORS

It is useful to consider sensors with respect to the type of radiation they deal with.

Figure 3.1 shows the spectrum of electromagnetic radiation, and in the upper half indicates the operational range of some 'conventional' sensors. In the lower half of the figure is shown the range of operation of some of the satellite systems that have been employed to date. It can be seen that the majority of satellite-borne sensors operate in the visible to near-infrared region, with some thermal scanners extending into the thermal infrared region. Some attempts have been made at using microwave sensors such as the Synthetic Aperture Radar (SAR) carried on Seasat, and radar sensors will certainly be important in the 1990's e.g. with the European ERS-1 satellite. Radar will also continue to be important with regard to air-borne sensors. For this study, though, the important sensors are in the visible and infrared, which are those which have been used routinely to provide a commercial service, and for which there is an extensive archive. These are the sensors carried by the Landsat and SPOT family of satellites, and their area of operation is shown in Fig. 3.2. Similar sensors are carried in the Indian and Russian satellites, but the archive of data is very small in comparison. Other satellite platforms carry sensors operating in these regions, and have an archive of data, but they offer a much coarser resolution and are used primarily for meteorological applications, not earth observation e.g. Meteosat, TIROS.

3.2 SENSOR PLATFORMS

The earth-observation sensors are carried on platforms which are in a polar orbit at an altitude of about 800kms. Orbits are arranged so that the satellites are sun-synchronous i.e. they always pass over any point on the ground at the same local time (usually mid morning) so that lighting conditions remain as similar as possible for images of the same area on different dates. The angle of view of the scanning system and the altitude of the platform determine the width of the strip on the ground which is actually recorded as an image (Fig. 3.3).

Using Landsat 4 as an example, the satellite has an altitude of 705km, and an orbit which is inclined at an angle of 98.2 to the lines of longitude. The swath width is 185km, and successive orbits are 2100km apart. The same spot is reviewed once every 16 days, and viewing time is 9.45am local time.

By comparison, the SPOT satellite is at an altitude of 822km, and gives a swath width of 60km. Its repeat cycle is 26 days, and viewing time is 10.30am at the equator. The principle difference from Landsat is that SPOT incorporates an orientable mirror, which can be pointed so that a swath is viewed which is not immediately below the satellite. Thus points up to 475km from the nadir may be viewed, and stereo pairs can be generated. More importantly, the movable mirror gives opportunity for a far greater viewing frequency than the 26 days dictated by orbital considerations.

The satellites such as Meteosat are in a geo-synchronous orbit i.e. they remain fixed relative to any point on the earth's surface. Geo-synchronous orbits are equatorial, and are at an altitude of 35,000km. A complete hemi-sphere is viewed at all times, but from such a great altitude that the resolution offered by these sensors is very coarse. However, these systems offer a very frequent transmission of images (one every 20 minutes), and the data may be captured using low-cost receiving stations. Thus, they are invaluable tools in the study of hydrology e.g. for rainfall estimation and regional vegetation monitoring, even though the data are too coarse for catchment studies. Similar comments apply to the polar-orbiting meteorological satellites such as the NOAA-TIROS family.

3.3 DATA TRANSMISSION

Satellite sensors generate a vast amount of information, at a very great rate. For instance, the Landsat Thematic Mapper acquires data at the rate of 110 megabits per second, and needs 10^{14} bits to give complete coverage of the whole of the globe (Hardy, 1985).

If a ground receiving station is nearby, then data can be transmitted virtually as they are collected. If not, then other strategies must be used. The earlier Landsat satellites had on-board tape recorders which could hold data until there was opportunity to transmit to a receiving station some time later. Landsat 4 and 5, however, depend on the relay of data via a communications satellite, thence on to the main receiving station in USA.

SPOT data is normally transmitted directly to ground receiving stations when activated, but there are also facilities for on-board recording of data, which is transmitted later to the Mission Centre at Toulouse.

Distribution of imagery is now handled by commercial companies so it is possible to purchase images in either digital format (computer compatible tape) or as a photo-product through these companies, or alternatively from their agencies. Normally these images are processed to make certain radiometric and geometric corrections. Radiometric corrections take into account the calibration factors for the sensors, and the optical and telemetry systems. Geometric corrections are applied to remove errors associated with viewing angle, satellite jitter, rotation and curvature of the Earth's surface. Thus, gross inaccuracies are removed, but much work is left to the purchaser to improve the clarity of the images, and to warp them to fit the geographical coordinate system of interest.

3.4 IMAGE PROCESSING

Data, when it is purchased is normally in the form of "images". Each image comprises a 2-dimensional matrix of Digital Numbers for each Band in which the sensor operates. Thus, for instance, Landsat TM data comprises 7 matrices, while SPOT HRV data comprises 3 matrices. The number of cells in each matrix depends on the ground area represented by each cell, and on the swath width of the sensor. Thus each matrix in a TM image has a nominal size of 6468 x 5375, and each picture cell (pixel) represents 30m x 30m on the ground, giving the 185km swath width.

The Digital Number in each cell represents the intensity level of the electromagnetic radiation received from the corresponding area of the earth's surface. Raw data are usually in the range of 6-7 bits, but conventionally, image data are represented by 8 bits, i.e. in the range 1-256.

Thus the image may be structured as shown in Fig. 3.4 or alternatively it could be structured as just one N-dimensional matrix. The way in which the Digital Numbers are ordered on the magnetic tape holding the data will depend on structure which is chosen.

The total of Digital Numbers in an image can be very large e.g. 35×10^6 numbers for each band of a TM scene, requiring a total of 243 Mbytes of computer memory for the whole scene. Usually, a subsample of the data is viewed, to assist in the selection of a portion of the image which is of interest to the job in hand, and to choose bands which contain the most useful information. These data are then transferred to a dedicated image-processor or work-station where image enhancement and interpretation may be carried out interactively. These vary in available work-space from 512 x 512 pixels in 3 bands up to 1024 x 1024 in 6 or 7 bands.

The image processor facilitates the visual display of data by illuminating screen pixels, corresponding to image pixels, at a level of brightness dictated by the Digital Number. The image processor thus provides a pictorial representation in grey-levels, of the degree of reflectance from the corresponding portions of the earth's surface. A visual interpretation of the reflectance in each spectral band offered by the sensors can be made, and, understanding the physical processes which determine the degree of reflectance, conclusions can be drawn regarding the nature and condition of the surface shown in the image.

If the data were panchromatic i.e. covering the whole of the visible spectrum, then the interpretation of the image would follow the same lines as an interpretation of an aerial photograph. Data are not normally panchromatic, but relate to specific regions of the electromagnetic spectrum. Thus it is only when viewing a single band in grey-tones that air-photo techniques are important. In most other cases a quite different strategy is adopted.

Whatever the strategy, there is a certain amount of processing which has to be done before the raw image is useful. This is concerned with the range of the digital numbers, and with their geometrical accuracy. Firstly, the sensors are equipped to deal with a wide range of reflectance levels, from Arctic snow to darkest jungle, whereas in any one particular scene, the range is far smaller. Thus if the raw data were displayed, all parts of the scene would have similar brightness and it would appear uniformly grey. An essential part of image processing is therefore to broaden the range of Digital Numbers in order to enhance the contrast in the image. This process is known as contrast stretching, and is illustrated in Fig. 3.5. This transformation is carried out frequently and may be performed simply by modifying the look-up tables of the display unit.

Secondly, although gross systematic geometric errors have been corrected by the supplier of data in their pre-processing, the geometry of the image is still not wholly precise and smaller, random distortions are still present in the image. Commonly, the user wishes to make the image conform to some established format e.g. he wishes to overlay the image on another image, or onto a conventional map coordinate system. A suitable transformation of pixel locations is achieved using ground control points, which are precise locations clearly identifiable on both the image and the map. The image x,y coordinates and the map x,y coordinates for each control point are established, and then a pair of transformation equations are calculated which will transform all other image coordinates to their appropriate map coordinates. In effect the whole image is resampled. Unfortunately, this can lead to a loss of information concerning radiometric values for the original locations but, in most practical applications geometric rectitude is essential, and the loss of radiometric precision has to be accepted. Geometric correction is very demanding in computation, and would normally be done using the host computer, although increasingly it is carried out in dedicated hardware at the work-station.

3.5 IMAGE INTERPRETATION

The simplest approach to image interpretation is to view the image of a single band in grey-tones, and to consider the tone, texture and context, in a similar way to interpreting an aerial photograph. This would be done bearing in mind the spectral signature of common types of surface cover as indicated in Fig. 3.6. A popular and cheap approach to image interpretation is based on the use of photo-prints of images in the near infrared region, which can be seen from Fig. 3.7(a) to give good separation of the major land-cover types.

Since the human eye detects differences in colour far more easily than it detexts differences in grey levels, single band images are often colour coded to indicate brightness levels (density slicing). An example of density slicing employed to examine variations in infrared reflectance from a reservoir is shown in Fig. 3.7(b). Just one density slice can be used to great effect in highlighting features of interest, such as the bodies of surface water shown in Fig. 3.7(c).

The visual approach can be developed by using several wavebands in combination. For instance a simple ratio such as Band 4/Band 6 can produce a fresh set of Digital Numbers to be displayed as grey-tones or as a density slice. The interpretation of the values of these numbers may become difficult, but the use of ratios does give the important benefit of removing the effects of shading (i.e. topographic effects), since the degree of shading will be the same in every band.

Alternatively, it is possible to assign one band to each of the colour guns of the display e.g. Band 4 to red, Band 6 to green, Band 7 to blue.

This produces a "false-colour composite" e.g. Fig. 3.7(d), in which the various colours may be interpreted in relation to the spectral signatures of the land-cover types. Various band/colour combinations have become standard for the various sensors, and much experience has been developed in their interpretation.

In the case of Landsat TM there are seven bands to choose from, and the various combinations for ratios and false colour composites becomes bewildering. In this case, it is possible to compress the data by a transformation of axes so that new band axes are selected which contain more information than any of the original axes. These techniques (e.g. principal component transformation, canonical analysis) produce very efficient data sets, but are mistrusted by many interpreters because of the difficulty of assigning any physical significance to values of the Digital Numbers (e.g. Fig. 3.7(e)).

A technique which overcomes this problem, while remaining objective, is that of classification. In this operation, the a priori knowledge of the image area is used. Groups of pixels representing areas having the same ground cover are identified within the data set. The spectral properties of these pixels are noted, and a region within n-dimensional space is demarcated wherein all pixels in the training areas lie. The procedure is repeated for other types of land-cover, and fresh boundaries are calculated. Eventually the n-dimensional space represented by the n bands of the data has been mapped with regions corresponding to the chosen land-cover types. The whole image is then reviewed and each pixel is allocated a land-cover according to its position in the n-dimensional space. The classifications can then be colour coded to produce a land-cover map that has been decided objectively, not by subjective, visual interpretation (Fig. 1.3.7(f).

The simple case of a 2-dimensional space is shown in Fig. 3.8. In this example pixel A would be classified as soil, pixel B as water, while pixel C would be unclassified. It may be noted from this simple illustration that the technique of density slicing would not provide a classification of these two land-covers because there is overlap in both bands i.e. the land-cover types could not be distinguished by their spectral response in Band x alone nor Band y alone.

The foregoing discussions apply to the interpretation of an image in order to distinguish certain areas within the scene e.g. to distinguish areas of healthy crop from diseased crop, or to map areas of different land use. The techniques noted there are not appropriate for the interpretation of linear or cultural features. Such features are normally interpreted visually following enhancement of the features by the use of filters (although some algorithms for objective line detection have been developed and are discussed later). The simplest enhancements are made using box filters of various sorts, (i.e. in the spectral domain) but it is also common to filter in the frequency domain having first made a Fourier transform of the image.

An important technique to emerge in recent years is that of contextual analysis, in which pixels are examined not individually, in isolation from the rest of the image, but in relation to neighbouring pixels, both near and far. In this way interpretations can be made which retain the spatial integrity of the data.

Bibliography for fundamentals of remote sensing

1. R.N. Colwell (Ed) 1983. Manual of RS (2nd Edn) Am.Soc. Photogram. and RS, Falls Church, VA.

2. P.J. Curran, 1985. Principles of Remote Sensing, Longman, London.

3. F.F. Sabins, 1986. Remote Sensing Principles and Interpretation (2nd Edn), W.H. Freeman, New York.

4. E.C. Barrett and L.F. Curtis, 1982. Introduction to Environmental Remote Sensing (2nd Edn), Chapman and Hall, London.

5. Landsat Data User Notes, Issues 1-35, Pub. NOAA Landsat Customer Services, Sioux Falls, SD, USA.

6. EOSAT Landsat Data Users Notes, Pub. EOSAT, Lanham, VA, USA.
SPOT Newsletter, Pub. SPOT-IMAGE, Toulouse, France.

7. National Remote Sensing Centre, 1985, Data Users Guide, NRSC, Farnborough.

PART 1 CHAPTER 4 A REVIEW OF HYDROLOGIC MODELLING.

Hydrology is the science of naturally occurring water, including its occurrence, circulation and distribution, and its chemical and physical properties. It is the science which underpins the essentially practical activities of water resources development and management. Hydrologic modelling is the body of mathematical methods whereby hydrologic processes are represented in such a way that conclusions may be drawn which are useful in water resources.

The science of hydrology has a venerable history, reaching back perhaps 5000 years to the early Egyptian civilizations, which were highly dependent, as now, on the annual flood of the Nile (Biswas, 1971). It is a moving experience for the hydrologist to contemplate the beautiful Roda Nileometer, and to reflect on the uninterrupted record of river levels stretching from 641AD up to the time of completion of the Aswan Dam. The record, though, was used only to estimate the timing and magnitude of the Nile flood, and little interest was shown in understanding the mechanism that caused it. It was not until the 17th century that Perrault and Mariotte carried out measurements relating rainfall to streamflow, and began to put hydrology on its modern footing.

In the 18th and 19th century much essential theory was developed, concerning evaporation (Halley, Franklin, Dalton), rainfall (Herberden) and streamflow (Chezy, Venturi). Crucially, in the early part of the 19th century the seeds of modern calculating capability were sown (Fleming, 1975). Babbage presented his plans for his difference engine, which was to be developed by Scheutz, to emerge finally in 1939 at Harvard as the digital computer.

But before computing power was freely available simple numerical representations of hydrologic parameters began to emerge. Mulvaney in 1851 presented the now famous (or infamous) Rational Formula Q = CIA, introducing the concept of "time of concentration", and in 1865 Dickens proposed the important idea of the envelope of floods, given by

$$Q = c\ A^n$$

which in various guises has retained favour, right up to modern times.

These, though were essentially empirical formula which, without probablistic statements to accompany them were of limited use. It was not until 1914 that Hazen began to lay the foundation for the statistical analysis of hydrological records (Hazen 1914,1930) which enabled statements to be made regarding the probability of particular flow rates. The statistical treatment of streamflow was given its contemporary character by Gumbel, who in a series of papers (e.g. Gumbel 1941) developed a theory whereby the probability of extreme events, such as the floods and droughts of such importance in water resource planning, could be evaluated (Gumbel 1958).

The empirical approach has persisted in the estimation of flood flows, particularly in the approach where some parameter of the probability distribution of floods is related through regression analysis to physical features of the catchment. Nash, to whom much of the modern empirical approach is owed, proposed a method in which the mean annual flood was estimated from catchment area and rainfall or slope, while the relationship between the mean annual flood and more severe events was estimated from a regression on the same characteristics (Nash and Shaw 1965).

Nash and Shaw used all available catchments in Great Britain with records of 6 years or more (57 in all), and used a linear regression analysis to find mean annual flood $\overline{Q}$, and its coefficient of variation, in terms of catchment area (A), slope (S) and mean annual rainfall (R). The best fit equations were alternatively

$$\overline{Q} = 0.009\ A^{0.85}\ R^{2.2} \qquad \text{(Imperial units)}$$

or

$$\overline{Q} = 0.074\ A^{0.75}\ S \qquad \text{(Imperial units)}$$

Mean annual rainfall was thought to be a poor substitute for some measure more directly linked with flood events, but was used as a matter of expedience. The coefficients of correlation for the log-transformed data were 0.92 and 0.87 respectively, which were remarkably high considering the quality of the data.

They further assumed that flood probabilities would follow either Gumbel's Extreme Value Type 1 distribution or the log-Normal distribution, and that the parameters of the distribution could be calculated from the mean and coefficient of variation of the flood series as estimated from the regression equations.

Cole, in the same symposium, proposed a method based on the simple empirical relationship

$$\overline{Q} = c\ AREA^{n}$$

in which the exponent n was fixed at 0.85 but the coefficient c

varied according to the geographic location of the catchment. These ideas have, together, matured into the method currently recommended by the Institute of Hydrology (NERC 1975) for the estimation of flood flows from ungauged catchments. In this method the mean annual flood $\overline{Q}$ is estimated from a set of catchment characteristics depending on location e.g. for East Anglia:

$$\overline{Q} = 0.302 \ \text{AREA}^{0.70} \ \text{STMFQR}^{0.52} \ (1 + \text{URBAN})^{2.5}$$

The T-year return period flood is then estimated using a growth curve which gives a value of $Q_T/\overline{Q}$ for each of nine hydrological regions. A similar approach, combining probablistic ideas and regionalisation, has been made for low-flows, (NERC, 1980), and has been adopted in many parts of the world other than U.K. (e.g. Drayton et al 1980).

In cases where a short hydrological record is available the Flood Studies Report recommends estimating the mean annual flood from the data, and then using the empirical relationship for the growth curve for the appropriate region i.e. a combination of statistical and empirical approaches. Only where an acceptable length of record is available does the Flood Studies Report recommend the use of Gumbel's Extreme Value theory to estimate the flood magnitude of interest.

Thus, the development of the empirical method of streamflow estimation can be traced from the mid 19th century to the present day. There was a branch point, however, in the early part of this century, following the development of the statistical theory by Hazen (1914), at which point probablistic hydrology and stochastic hydrology began to develop separately. Probablistic hydrology, as outlined above, is concerned with the probability that an event (flood or drought) will be equalled or exceeded. In stochastic hydrology the sequence of events in time is the all-important consideration. Such sequences are required in the design of reservoirs, and the determination of operational procedures i.e. they are the inputs for systems analysis.

Early attempts at generating synthetic flow records with the same statistical features as observed records were crude, and depended for random number generation on techniques similar to those seen at the gaming tables. These "Monte Carlo" methods were quickly made obsolete with the advent of electronic computers, and modern theory stems principally from advances made in the 1960's. Fiering, importantly, proposed the use of Markov chains to preserve high-frequency fluctuations in simulated streamflow, and developed simulation models further to preserve longer term cyclicity, seasonality and correlation with other streamflow records (Thomas and Fiering 1962, Hufschmidt and Fiering 1962, Fiering 1967).

These models are called "auto-regressive" in that they depend on the correlation between discharge at time t and at time t-1. A typical annual flow model is :

$$Q_t = \overline{Q} + \rho\,(Q_{t-1} - \overline{Q}) + \sigma(1 - \rho^2)^{\frac{1}{2}}\,\eta_{t-1}$$

i.e. the flow at time t is estimated from the mean annual flow, plus a proportion of the excess or deficit in the previous year, plus a random component scaled to give the appropriate variance. Such a model can be used to generate a long synthetic series, by generating random numbers from the appropriate probability distribution.

Seasonality can be built into an auto-regressive model, for instance the Fiering single-station seasonal model of monthly discharges, which preserves means and variances of monthly flow statistics.

$$Q_{t+1} = \overline{Q}_{\tau+1} + \frac{\sigma_{\tau+1}}{\sigma_{\tau}} \rho (Q_t - \overline{Q}_{\tau}) + (1 - \rho^2)^{\frac{1}{2}} \sigma_{\tau+1} \eta_{t+1}$$

The problem with the Markov process (i.e. lag-one auto-regressive process) is that it does not model long-period fluctuations such as were observed by Hurst (1951) in the Nile record and other long natural series. Streamflow models preserving both low and high frequency fluctuations were proposed by Mandelbrot and Wallis (1969), Mejia, Rodriguez and Dawdy (1972) and Box and Jenkins (1970). Whilst there is still controversy over the Hurst phenomenon, it is clear that stochastic models should be capable of generating realistic drought sequences. Such models have been developed to a highly sophisticated level by O'Connell (1974), Lawrance and Kottegoda (1977) and Lettenmaier and Burges (1977).

Later development was in the field of stochastic models which preserved the statistical relationships between several sites. Matalas (1967) developed a multi-site model for annual flows based on the simple first-order auto-regression model, which in matrix terms becomes

$$Q_t = X_t S^T + \overline{Q}$$

where

$$X_t = A X_{t-1} + B E_t$$

where the X terms are vectors of the series of standardized flow values at each site.

Developing a multi-site model to generate monthly values requires the estimation of very many parameters, and creates potentially difficult problems in the matrix algebra. An alternative approach favoured by Lane (1979) is to generate multi-site annual data, and to disaggregate them to monthly or seasonal values.

The fundamental flaw in stochastic modelling is that it does not represent the physical processes involved in streamflow generation : it merely sets out to replicate statistical properties. (Although it must be said that certain models such as the first-order auto-regressive model reflect the storage of precipitation as groundwater, and its subsequent slow release). So another major strand in the development of hydrology has to be considered; that in which all the processes within the hydrological cycle are described mathematically, and linked together so that the response of a catchment to climatic inputs can be simulated. Such models are known as deterministic models, and come in various guises, suitable for the different demands placed on them. Linsey, Kohler and Paulus (1982) distinguish research models, conceptual models, lumped models, distributed models, continuous models and event models.

Research models use formal mathematical expressions for the algorithms used to evaluate changes in the hydrological processes. Solution of many differential equations is laborious, and these models are normally used to examine the effect of subtle changes, or to examine processes over very short time periods.

In conceptual models the algorithms are often simplified by the use of empirical relationships to speed computation and make large models feasible, whilst retaining their ability to simulate in detail over time and space. Most models used in water resources are of this type. Another common method of simplifying models is to consider large sections of the catchments and their climatic inputs as if they were homogeneous, and to lump them together. Distributed models, on the other hand, consider the catchment as a large number of sub-areas, each modelled separately, and then combined together to simulate the whole catchment response.

Continuous models aim to provide simulation of streamflow.over many years, and carefully maintain a water-balance throughout so that the catchment condition is always known. Event models aim at simulating just one flood hydrograph for a particular rainfall imput, and require that initial catchment conditions are specified.

In practice, the variations are quite limited. Most common are conceptual continuous models, usually lumped to some degree, which are used to simulate streamflow sequences for systems design, or are used to predict streamflow for real-time operational decision-making (e.g. reservoir operation, flood warning). Alternatively, there are event models, usually lumped, which are used to predict flood flows in the design of drainage works.

Deterministic modelling depended on the development of many theories concerning the processes in the hydrological cycle, and equally depended on improvements in climatic measurements and streamflow gauging to produce the high quality records required in the construction and calibration of models. In the former, the seminal work of Penman (1948,1956) and Thornthwaite (1944) in evaporation, Horton (1933) in the run-off process, Theis (1935) in groundwater movement, and Lighthill and Witham (1955) in kinematic wave theory should be mentioned. In the latter, credit must be given to the US Geological Survey development of field survey technique, and to the World Meteorological Organization for defining good practice in all hydrologic field work (World Meteorological Organization 1974).

Above all else, though, it was the coming together of Crawford and Linsley at Stanford University that put deterministic modelling on its present footing. Their collaboration resulted in the formulation of the Stanford Watershed Model (Crawford and Linsley 1960, 1966) which has developed over many years, and has formed the basis for many other continuous conceptual models. Perhaps even more important was the collaboration between Linsley, Kohler and Paulus, who brought together in a usable form the ideas from the many disciplines which are used in the study of hydrology (Linsley, Kohler and Paulus 1949, 1982).

At a simpler level we have the lumped models, in which a rainfall event is modified by some operator to produce a model of the runoff event. In practice the operator is often defined empirically, and referred to as a 'black-box' i.e. this is a pragmatic approach in which no attempt is made at understanding the process by which runoff is generated. The rainfall event may be very long-term e.g. annual rainfall, in which case a simple model of catchment yield is given by the relationship :

Annual Yield = Annual Rainfall - Losses

where it is assumed that the net groundwater change over the year is zero, and that losses are entirely evapo-transpirative. Alternatively a short-term event model may be proposed i.e. a flood model of the form

$$Q_t = f\ (\text{Rainfall})$$

where the flood discharge at time t is some function of the rainfall. This is often approached in two stages : firstly an estimate is made of the total volume of runoff and secondly that volume is distributed through time.

In Britain, the estimation of flood runoff has traditionally been done by the use of a runoff coefficient, whose magnitude was determined by the experience of the hydrologist. In the USA a very popular method was developed by the Soil Conservation Services (USDA 1962) in which the volume of runoff was determined by a coefficient (or Curve Number) which depends on the land-cover and type of soil in the catchment i.e. an empirical relationship.

Following the estimation of runoff volume, the flood magnitudes over a period of time are provided by scaling up a Unit Hydrograph i.e. the flood hydrograph resulting from a unit of effective rainfall uniformly distributed over the catchment in a unit time. Models for Unit Hydrographs have been developed by Snyder (1938) and Nash (1960) and many others, in which the shape and proportion of the hydrograph are related empirically to the size and shape of the catchment. Such models formed the basis for the bulk of engineering hydrology since the 1960's, and have been modified (and simplified) by the work of the Institute of Hydrology in the Flood Studies Report (NERC, 1975).

Since the 1960's there has been an explosion in the interest in hydrology, fuelled by the growth in institutional arrangements such as (in Britain) the Water Resources Act of 1963, the formation of the Institute of Hydrology in 1962 and the Water Resources Board in 1963. The availability of data increased markedly with the initiation of the International Hydrological Decade 1960-70, but most important in terms of modelling was the rapid improvement in the capabilities of the digital computer. Hydrologic modelling in its present form now appears to depend far more on computational ease rather than scientific understanding. Indeed, the roles are often reversed, and numerical modelling is used as a method for improving scientific understanding.

In summary then, modern techniques in hydrological modelling include :

1. Empirical relationships, which define statistical features such as mean annual flood,
2. Stochastic models, which generate long synthetic series with the same statistical properties as an observed series,
3. Conceptual models, which simulate the long term response of a catchment to climatic inputs, and
4. Event models, which predict the streamflow resulting from particular rainfall events.

At the risk of oversimplifying, it may be said that types 1 and 4 are favoured by engineering hydrologists, and those concerned with preliminary evaluations of the water resources of an area. Types 2 and 3 are employed only where individual projects are of sufficient importance to warrant the considerable effort, and where sufficient good quality data are available.

PART 1 CHAPTER 5 INTRODUCTION TO GEO-INFORMATION SYSTEMS AND DIGITAL TERRAIN MODELS

It has been noted above that hydrologic models may be "lumped" or "distributed". Lumped models are those in which land and climatic characteristics of the catchment are lumped into average parameters representative of the whole watershed, whereas, in distributed models, the catchment is treated as a number of elements or sub-areas of homogeneous character, similar in terms of soils, cover, topography and climate.

The benefit of distributed modelling is that the spatial and temporal variations in catchment characteristics may be investigated. The response of the catchment to non-uniform climatic inputs, especially rainfall, may also be analysed.

The problem in distributed modelling is acquiring and handling the very large quantities of data needed to build the models. From earlier comments it can be seen that remote sensing can play an important part in the acquisition of hydrologic data. Handling, managing and analysing the data is carried out using spatially organised data base systems known as Geo-Information Systems, or GIS.

Geo-Information Systems are claimed by Tomlinson (1984) to have originated in the 1960's as a means for national governments to make objective assessments of their natural resources. Perhaps the earliest system used was for Canada, and was aimed simply at making a spatial and temporal inventory of land resources. The development of GIS over subsequent years has been in managing such data systems so that modelling and simulation may be carried out.

The main functions of a GIS are :

- Data collection and input (maps, statistics, images)
- Data storage and retrieval (data base)
- Manipulation and analysis of data (modelling, simulation)
- Output of information (maps, reports)

The organisation of an idealised GIS is shown in Fig. 5.1 (from Young, 1986). The core of the system is a cartographic base, against which the location of data items may be recorded. On this base are placed a series of thematic overlays, that describe physical features such as drainage, soils, geology, land-cover etc, and socio-economic features such as population density, indicators of wealth etc. These latter have little relevance to hydrologic modelling, and it is the former category which are mostly encountered. A particular, simple, form of GIS which is frequently encountered is a model of the topography of an area, containing only information regarding the elevation of the ground-surface. These are known as Digital Elevation Models (DEM's) or when they include other basic information such as drainage and communications, they are known as Digital Terrain Models (DTM's).

When the approach to GIS work is via remote sensing, it is convenient to imagine the GIS as a series of overlayed images (Fig. 5.2), which all have the same geometric (or cartographic) basis i.e. a raster system. This leads to the idea of conceptualizing the hydrologic system as a catchment (defined by topographic properties) which may be divided into a large number of grid cells (usually square) which have a set of hydrologic characteristics (slope, soil, cover etc) and which are linked by location and by hydrologic processes such as run-off and interflow.

Alternatively, grid cells may be combined into sub-units which could be located as finite elements, which would be useful in surface run-off systems, but very much more so in groundwater models where the finite element approach is more traditional.

In summary, then, Geo-Information Systems provide a useful environment for the acquisition of disparate forms of hydrologic data, within which the data retain important spatial features and where models may be constructed to simulate hydrologic processes. Once constructed, the data sets may be easily up-dated using interpretations of satellite imagery, and given some rationalisation of spatial chracteristics the data may be blended with other sets of data and other types of model.

PART 1 CHAPTER 6 SUMMARY

In many parts of the world, the development and management of water resources is hindered by the lack of such basic information as inventories of surface resources, estimates of catchment yield, and the magnitude and frequency of floods and droughts.

Space technology has provided us with the opportunity to acquire information from orbiting satellites, ranging from the high altitude geo-stationary satellites such as Meteosat, which provide coarse resolution data in real time, to the SPOT satellites which provide high resolution data, but on a less frequent basis. For the preliminary assessment of the water resources of a region, there exists a large archive of data from the Landsat family of satellites, which provide data over a wide range of the spectrum at good resolution, and at modest cost. The data may be interpreted using subjective visual techniques, interactive computer techniques or automatic computer techniques to provide maps and statistics of surface features. These interpretations may be used in the estimation of water resources.

Techniques for the numeric modelling of streamflow are well established. For basic water resource estimation the simpler empirical approach is appropriate, and requires the least data. The demand for such data can be easily met from satellite systems. More sophisticated models, capable of defining spatial and temporal features of streamflow exist, but they have a very much greater demand for data. They may be served well by a combination of satellite sources providing real-time and historic data, especially when the data are handled in spatially organised data bases and are blended with data from a variety of other sources, such a digital terrain models.

Traditionally, empirical models have been based on a set of catchment characteristics, such as catchment area, drainage density, land cover, soil type and slope. They are often based on regions of similar hydrology, so that climatic factors do not enter the statistical model. Geomorphic factors such as catchment area and stream frequency, and thematic factors such as land-cover may be estimated from satellite data to provide timely inputs to models, but require fundamentally different techniques of image processing and interpretation.

The breaking down of environmental information into many separate categories is a basically inefficient procedure, and has evolved largely because of the restrictions which are imposed by the traditional methods of map-making. Satellite data is holistic, because the spectral signature of any point on the ground expresses the combined influence of all the factors at work there: the soil type and its state, the vegetation, the slope and aspect of the ground and so on. Satellite data should thus be more efficient than conventional map parameters when used in environmental models.

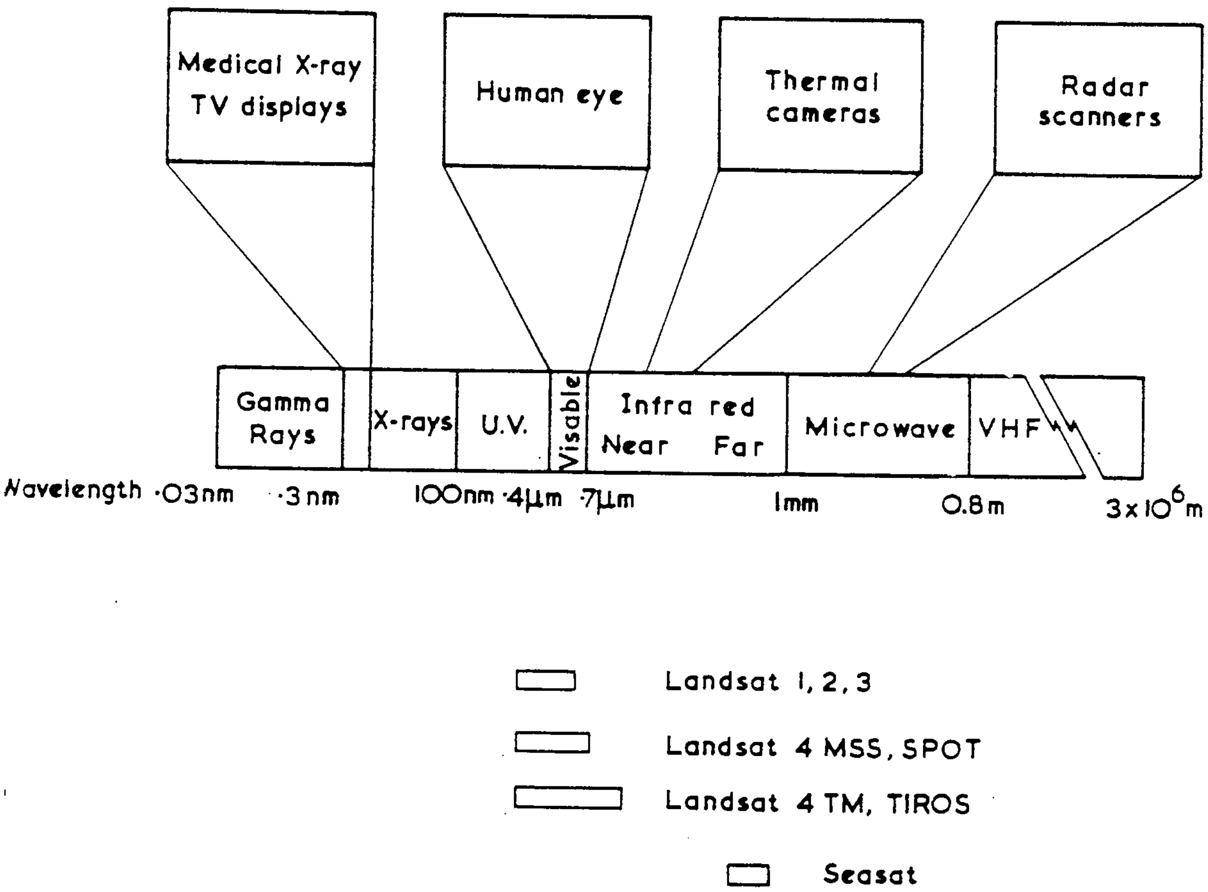


FIG: 3.1. SPECTRUM OF ELECTROMAGNETIC RADIATION.

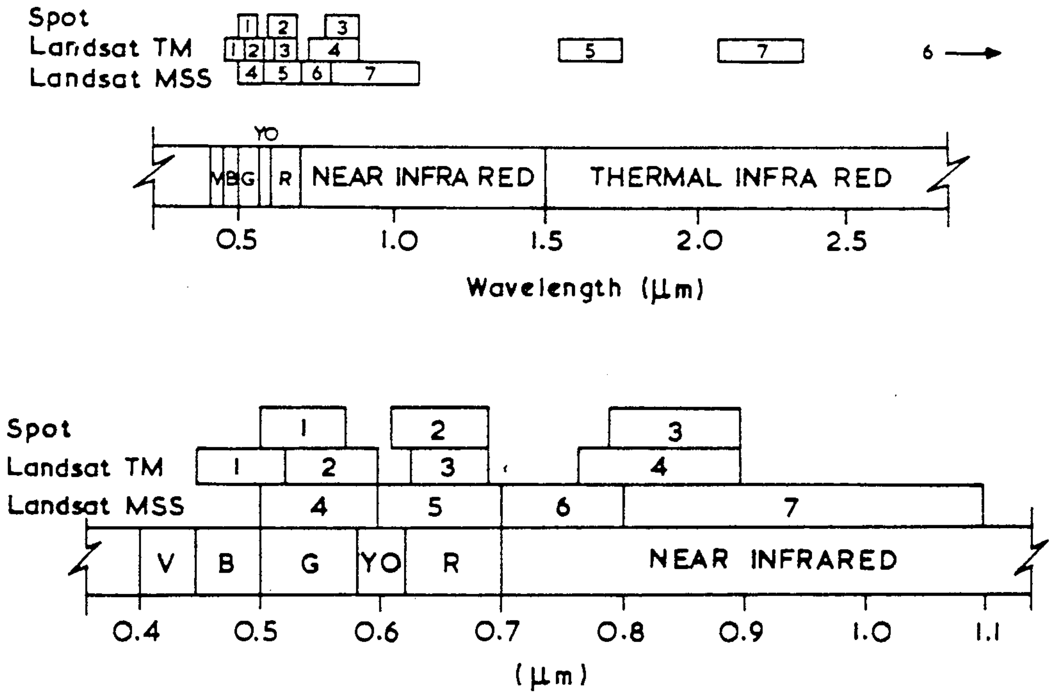
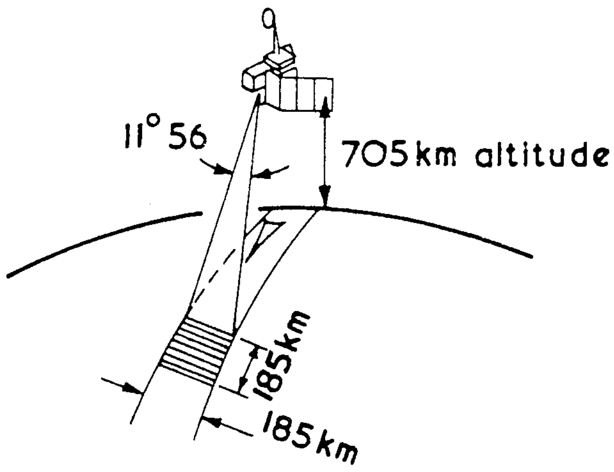
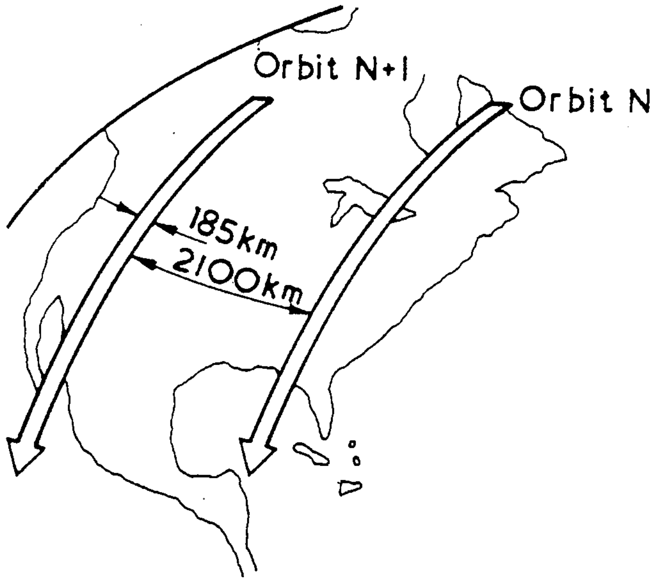


Fig. 3.2 ELECTROMAGNETIC SPECTRUM AND SENSOR WAVELENGTHS

(Note: There are no sensors in bands 1,2 or 3 of Landsat MSS)



SWATH PATTERN LANDSAT 4



ORBITING PATTERN LANDSAT 4

FIG: 3.3 ORBITS AND GROUND SWATHS.

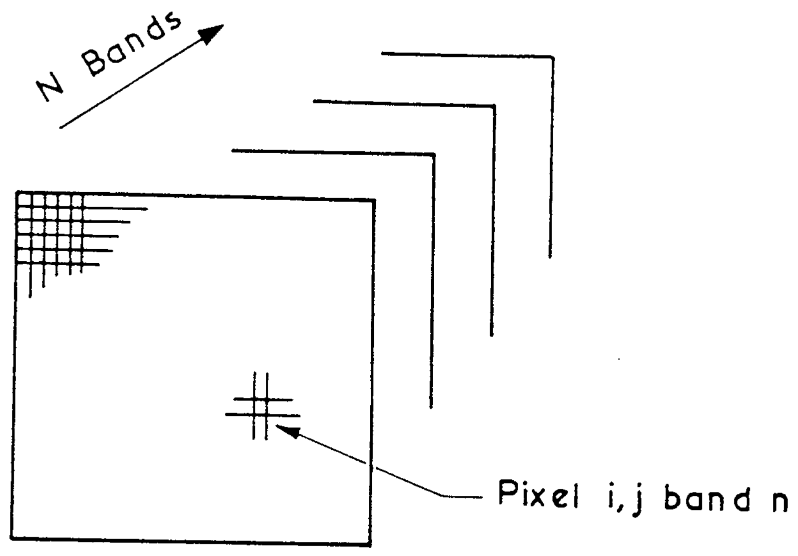


FIG: 3.4 STRUCTURE OF AN IMAGE IN DIGITAL FORMAT.

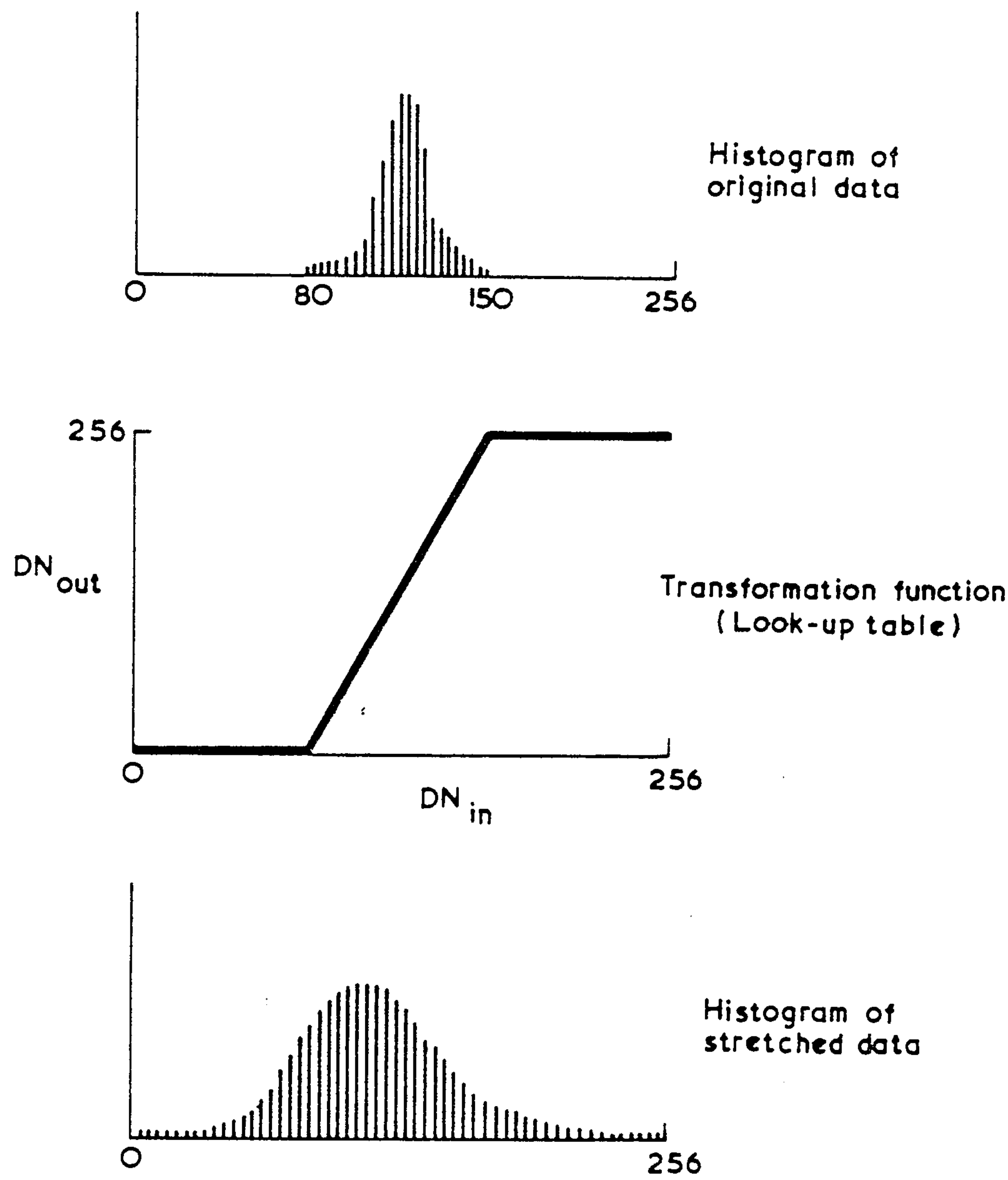


FIG: 3.5 CONTRAST STRETCHING

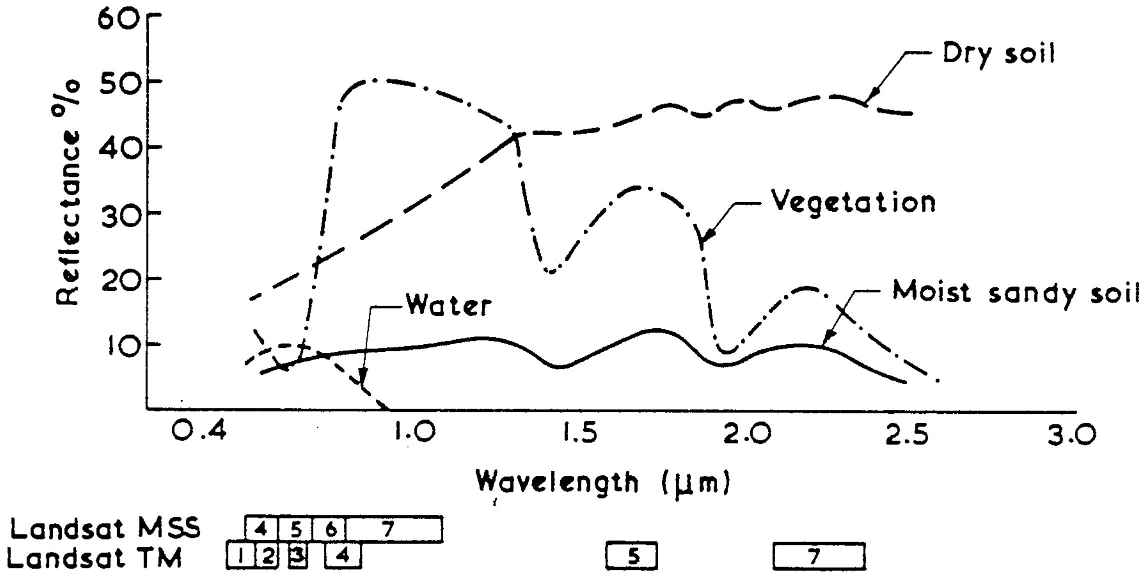


FIG: 3.6. SPECTRAL SIGNATURES.

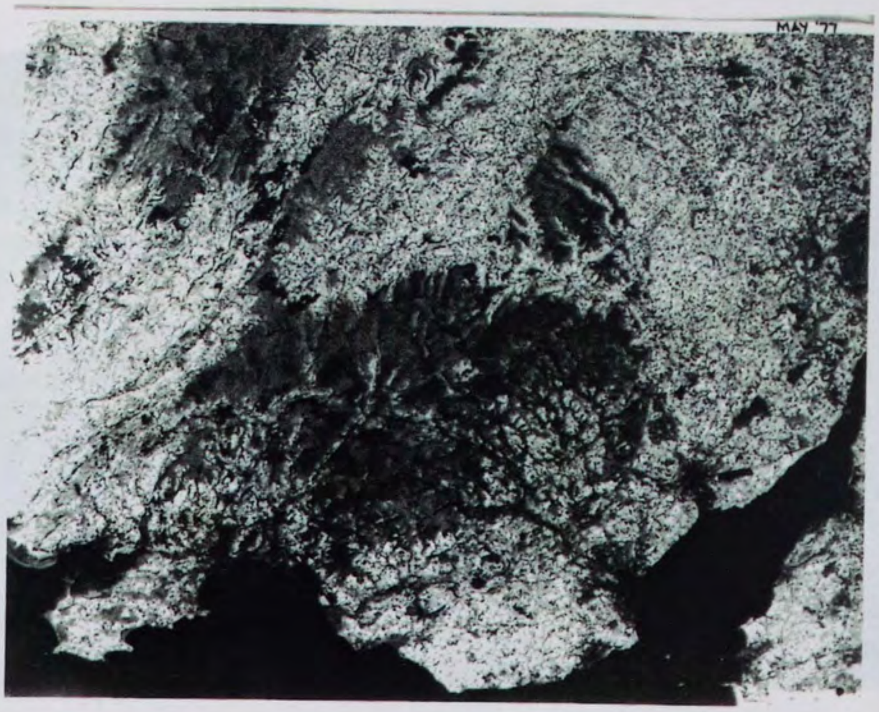


Fig. 3.7(a) Near Infra-red image

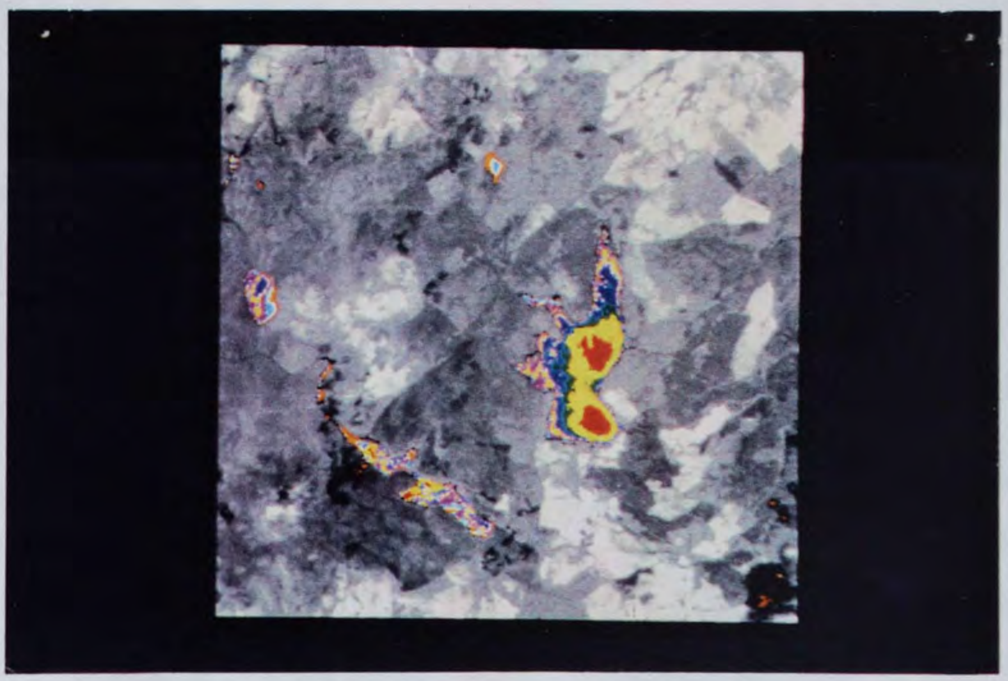


Fig. 3.7(b) Density slice

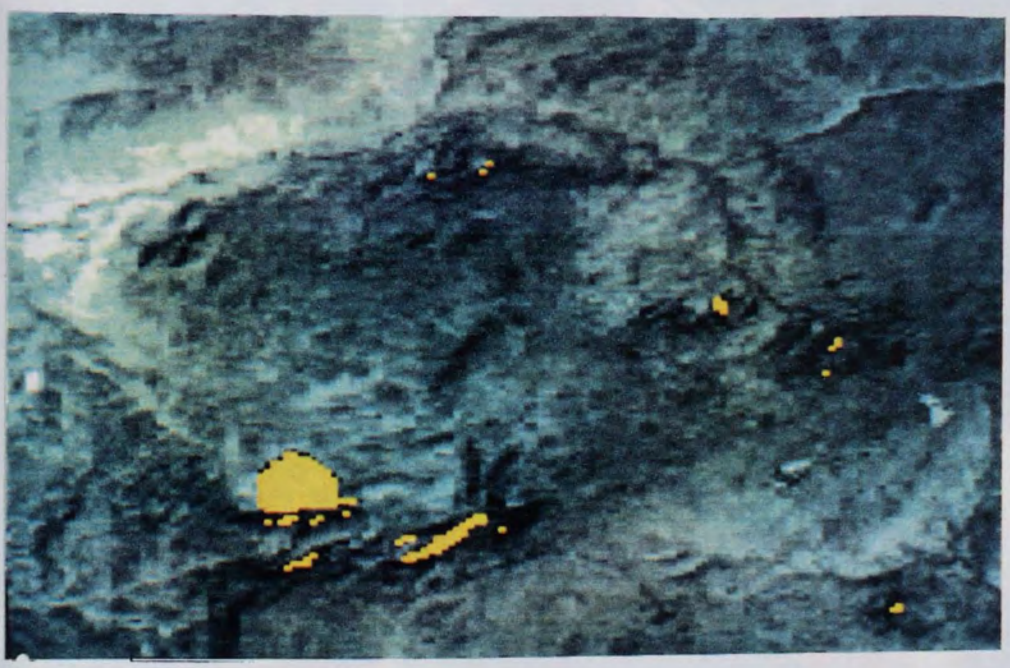


Fig. 3.7(c) Highlighting surface water

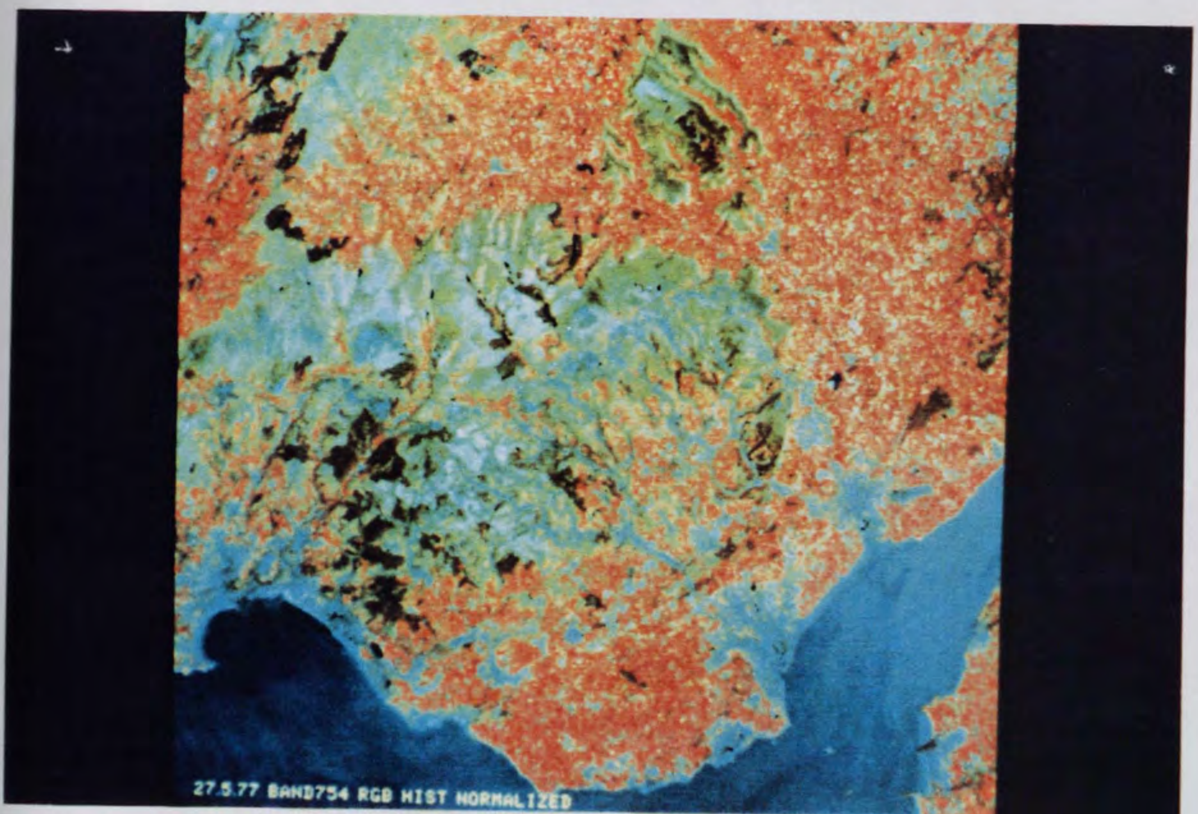


Fig. 3.7(d) False colour composite

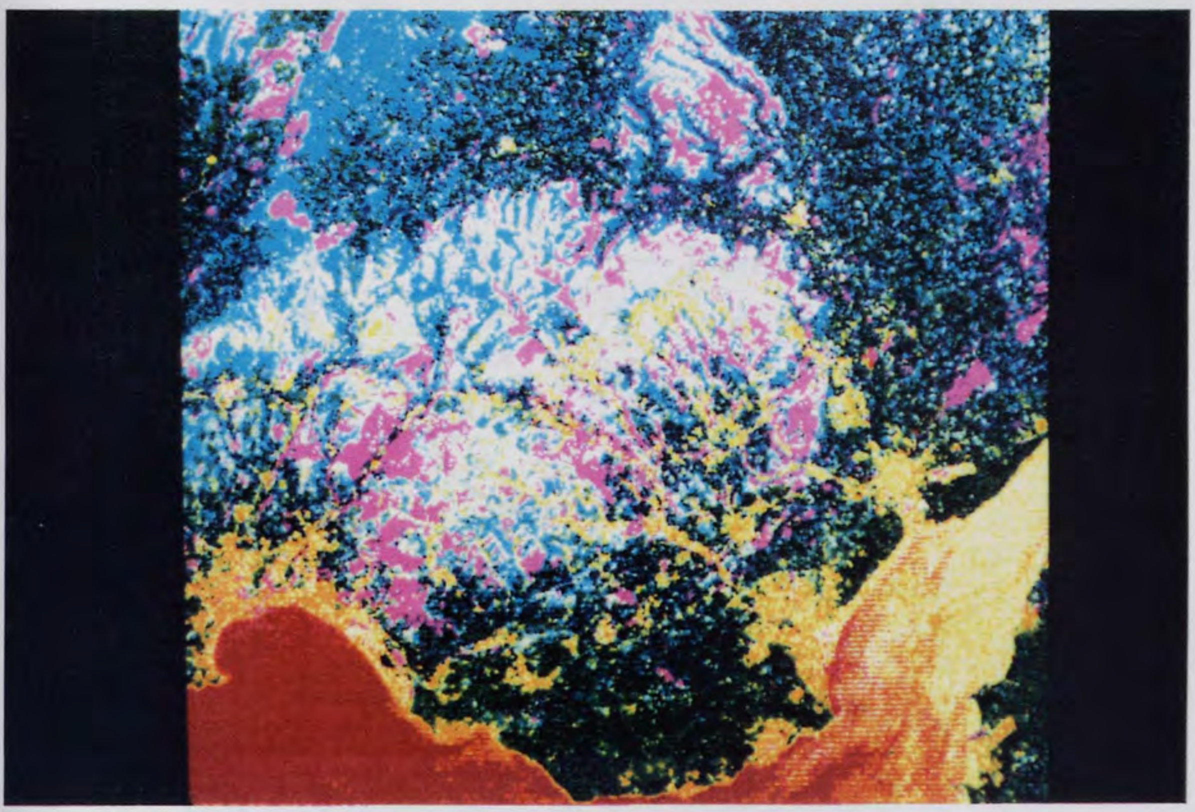


Fig. 3.7(e) Principal components

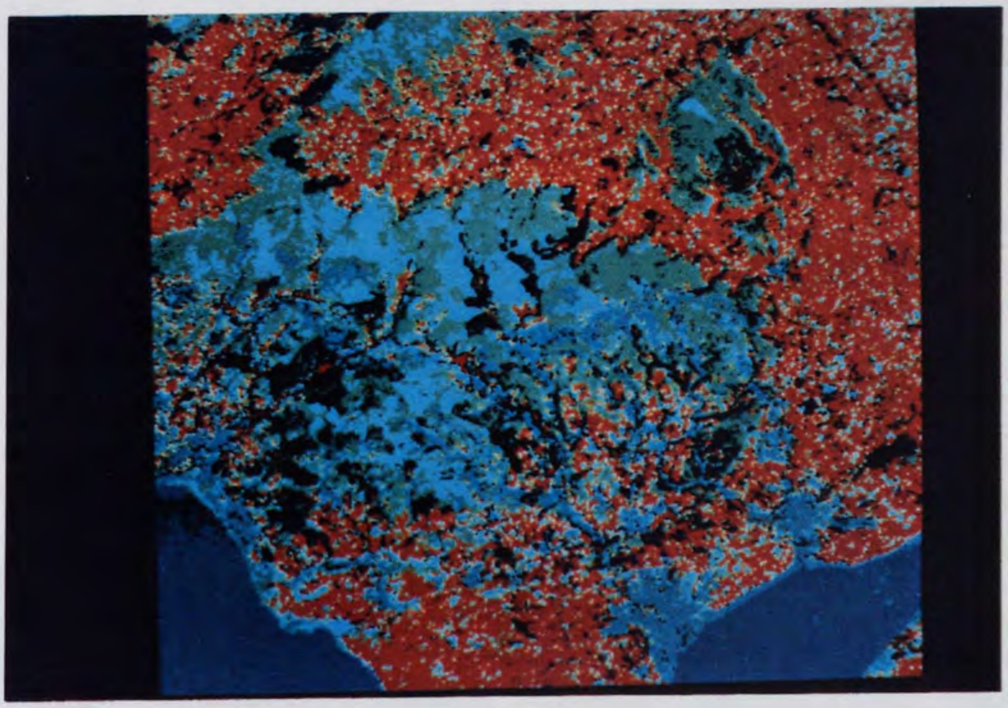


Fig. 3.7(f) Land-use classification

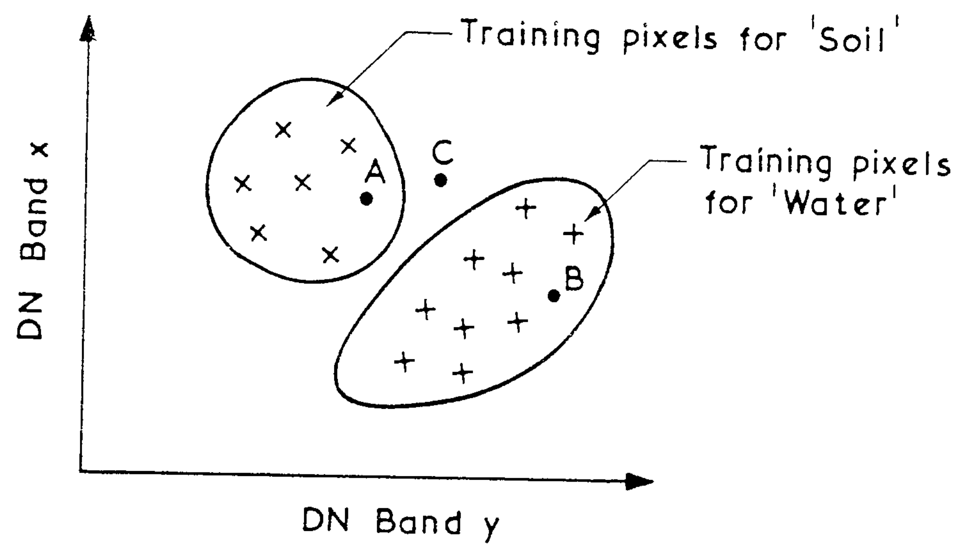


FIG: 3.8. Classification in 2-dimensions.

Aston University

Content has been removed for copyright reasons

Aston University

Content has been removed for copyright reasons

Fig. 5.1 Organisation of a GIS (from Young, 1986)

Aston University

Content has been removed for copyright reasons

Aston University

Content has been removed for copyright reasons

Fig. 5.2 Data layers of a GIS (from Curran, 1985)

PART 2 MAPPING SURFACE WATER RESOURCES USING REMOTELY SENSED DATA

PART 2

SYNOPSIS

A report is given of an investigation into the usefulness of satellite imagery for mapping surface water resources, concentrating on the mapping of drainage networks, and using visual techniques.

The need for new surface water maps is established in Chapter 1, following which a methodology is presented for the interpretation of drainage networks from satellite imagery. The difficulty of assessing the quality of maps is recognized and a methodology for making quantitative assessments is presented.

A review of published work in this area is given, and then three case studies are reported, firstly using visual interpretation of Landsat imagery for a study area in South Wales, then using higher resolution imagery, and finally using Landsat imagery for a semi-tropical region.

The success of the visual interpretation in providing maps of surface water, and in providing estimates of geomorphic parameters used in numeric models is examined and conclusions are drawn.

PART 2 CHAPTER 1

INTRODUCTION

1.1 The need for new hydrologic maps

Hydrologic mapping is needed when there are no maps of an area of interest or when existing maps cannot be trusted. Whereas most parts of the world have now been mapped at some level, it is often the case that it has been done at very small scales such as 1:1 million and 1:$^{1}/_{4}$ million, and that the resources needed for mapping at 1:50,000 scale and larger are not available.

Even where maps exist at a scale appropriate for water resource planning, features of interest such as rivers and lakes may not be represented on any logical or scientific basis. For instance, the Ordnance Survey of Great Britain has no formal criteria for deciding whether a stream should be shown, and the decision whether to include a surveyed stream in a map will be taken on grounds of cartographic clarity, rather than any criteria concerning the size or flow of the stream. (Personal communication, 1987). Consequently, the density of a stream network may change radically from one map-sheet to another. The lack of consistency from one scale to another was examined by calculating stream frequency for a number of study catchments from maps at 1:25,000, 1:50,000 and 1:250,000 scale, as shown in Fig. 1.1. It can be seen that there is some small degree of correlation between 25,000 and 50,000 scale maps (r^2 = 0.26), but that there is none at all between 25,000 and 250,000 scale maps. Remote sensing offers a way in which objectivity may be brought into the mapping of drainage networks.

Maps may also be simply out of date. In the parts of the world where large seasonal floods are experienced, river courses change dramatically from year to year, and satellite sensing offers a cost-effective way of monitoring those changes, as indicated in the Indus River study by Deutsch and Ruggles (1977). In Britain the details shown on 1:10,560 scale maps (from which smaller scales are derived) are taken from field work carried out in the 1950's and 1960's, and revisions are now only made in areas of urban growth, whereas many drainage networks have changed significantly in that period. The repetitive nature of satellite remote sensing offers opportunities for detecting and monitoring change over a period of time.

Hydrologic mapping may be carried out simply for the purpose of inventory, or may be used to provide information for hydrologic models. In the first instance, the requirement is to detect and locate all bodies of open surface water down to some prescribed size. This has been done successfully for lakes and ponds as small as 2.4 ha using Landsat MSS (NASA, 1977) and smaller than $^1/_2$ ha using SPOT data, as will be noted in 3.2 below. The changes in the size of lakes and swamps are often key issues in the management of a water resource. For example, the Okavango and the Sudd swamps, which are so crucial to African water resources, have been monitored using remote sensing by Hutton and Dincer (1979) and by FAO (1982). Rivers and streams also need to be mapped, so that the basic drainage system can be assessed, and they will need to be mapped over several years and seasons to distinguish ephemeral streams from persistent streams.

In modelling, the requirement is to provide values of the various geomorphic parameters which are relevant to the hydrologic process. Typically these include factors concerning the shape and size of the catchment area, the length of streams and the density of the drainage network (which may be defined by the presence of streams, or simply by the nature of the terrain). Whereas remote sensing is unlikely to be used to provide data for highly detailed conceptual models, it is ideally placed to provide data for the class of regression models based on catchment characteristics, exemplified by the work done by the Institute of Hydrology for the Flood studies Report (NERC,1975) and the Low Flow Studies (Institute of Hydrology, 1980).

1.2 Methodology - interpretation

On seeing the early Landsat images, interpreters were struck by the clarity with which open water was depicted. The reason for this clarity was that clear water reflects almost no radiation in the infrared, and very little in the visible regions (Fig. 3.6 Part 1), and so appears very dark in the image. Quite small bodies of surface water can be highlighted by the simple technique of density slicing an infrared image i.e. assigning a particular colour to all pixels whose value lies within some low range. This works well in the case of a pool lying in open countryside, as in Fig. 3.7(c) Part 1 , where a pond of just a few pixels size is clearly shown. It is far less successful in the case of streams and rivers, where the stream may be completely overhung by riverine vegetation, or where the width of open water is less than the resolution of the scanner (i.e. 80m for Landsat MSS down to 20m for SPOT). In those cases, it is likely that a combination of wet ground, shady riverine vegetation and some open water may give a reflectance level which is low (especially in the infrared), but not as low as that of an open lake. Streams may then be distinguished on the basis of dark tone, but may easily be confused with other dark-toned linear features such as roads, railways and canals.

The task of mapping streams is thus not a simple one. It cannot be done using a clear-cut objective crierion based solely on the data provided by the sensor, but becomes a process of interpretation, to be carried out by a human interpreter or by machine.

In the case of human interpretation we soon meet the problems of perception. Sabins (1987) recognizes the following factors which are important in the interpreter's ability to perceive and identify features:

Detectability : the ability of an imaging system to record the presence of an object.

Recognizability : the ability to identify an object in an image.

Signature : the spectral characteristics of an object.

Texture : the frequency of change, and arrangement of tones on an image.

Interpretation key : the combination of size, shape, tone and colour of an object.

In the case of rivers, texture is unimportant, and the signature is known (although not unique). The imaging system has the ability to detect quite small streams, which leaves recognizability as being wholly dependent on the interpretation key. The tone or colour is given by the signature, so we are left with size and shape as the key factors in the interpretation, which sadly are not easily defined for rivers. So, faced with an image, on which dark-toned linear features may be detected, how is the interpreter to decide which of those lineaments are in fact rivers?

The approach of a human interpreter is radically different from that of a machine-based system. The human interpreter brings to the task a wealth of knowledge of the world, and a very highly developed ability to distinguish and categorize shapes. Thus the human interpreter is likely to start the interpretation by viewing all of the scene, to achieve some understanding of the topography, recognizing geomorphic units within which the drainage system must fit logically. Moving from the general to the particular, he will distinguish rivers and streams on the basis of their characteristic shape. In well-dissected open terrain his task is relatively simple, but in areas of low relief, strong vegetation or modified drainage, he will need to adopt a rigorous set of objective criteria such as those required by a machine system. This objectivity is very necessary to overcome the psychological pressures on the interpreter. The human mind has a natural desire to create order out of chaos : faced with an image containing much fragmented and vague visual information, the interpreter will subconsciously want to "detect" lineaments that bring items together and form part of a logical structure. For example, the inclination for humans to see patterns of dots as continuous lines was demonstrated by Maunder, in 1913, in terms of drainage patterns, in a critical review of the work done by Lowell in plotting the canals on Mars from telescope observations (Baker, 1982). Looked at in a rather more simple way, the interpreter is given the job of drawing drainage lines, and naturally feels that the more lines he draws, the better he does the job. Thus objective criteria area needed to overcome these tendencies.

In machine interpretation the problem becomes one aspect of the general problem of pattern recognition. The most well advanced techniques are those for finger-print matching, and for deciphering writing on paper. In both these cases the solution can be arrived at by comparing the unknown to a series of templates until a good match is found. Unfortunately with rivers, there is no one shape or pattern that can be used as a template.

The task is thus broken down into a number of steps, the first of which is to detect all the lineaments in an image (not just rivers). This is normally done by using local operators, i.e. by observing the pattern within a 3x3 pixel window, and comparing it with a series of templates for simple lineaments in an attempt to make a match.

Having identified individual pixels which fit into a lineament pattern, then a scanning technique is used to link series of pixels into embryo lineaments. The difficult next step is to remove low-probability lines, and to make sensible joins between fragments to construct the full lineament map. Fundamental work in this area is reported by Gurney (1983), and the development of knowledge-based automatic systems is reported by Peacegood and Wilkinson (1985). An alternative, global approach is proposed by Bellavia and Elgy (1968) which is more appropriate for very noisy images.

Having identified linear features the problem remains of distinguishing between various kinds of lineaments such as roads, rivers etc. Wang et al (1983) propose a methodology for constructing a relational model in which a database stores line segments, with a list of properties for each segment, which a spatial reasoning algorithm uses to propose candidates for "streamhood". They make a useful list of criteria to be used in the decision process :

(a) completed pattern is dendritic in form

(b) completed network has no enclosed areas (i.e. stream segments are connected only at the downstream ends)

(c) completed network is without gaps (i.e. all segments are connected; no isolated segments)

(d) unconnected ends of the smallest tributaries point towards ridgelines

(e) tributaries form acute angles with main streams, right angles possible; obtuse angles unlikely

(f) streams and valleys cannot cross ridgelines

(g) ridgelines cannot cross streams

(h) junctions of 2,3, or 4 segments possible

(i) small tributaries to large streams typically occur parallel or nearly parallel to each other, and often join the larger stream at a right angle

(j) "larger" streams may be recognized by a larger radius of curvature

(k) "smaller" streams may be recognizable by a smaller radius of curvature

(l) each stream segment can be labelled with a flow direction such that each stream junction has exactly one outward flowing system

They recognize that the translation of these stream semantics into relational form will require a great deal of work. Nevertheless, the criteria listed here may be used to good effect by human interpreters, and visual interpretation is likely to remain the only practicable approach for some time. To quote Townshend (1981) :

"The power of the human interpreter with virtually random access to any part of the image, an ability to examine complex spatial image structures accompanied by access to a variety of sophisticated analogues of terrain....will all combine..to ensure the retention of visual physiographic methods for many applications".

Summarising, then, drainage networks may be mapped by the process of visual interpretation of images, in which drainage features are recognized as dark-toned lineaments possessing the typical geometric and topological characteristics of rivers. Computer-based interpretations are being developed, but are a long way from providing useful end products.

There are other approaches to creating drainage maps, which do not depend on the difficult process of recognition discussed above. For instance, in arid areas streams may be ephemeral, so that there is no characteristic reflection from water or wet soils, while in northern regions streams may be completely covered by ice and snow. In such cases an estimate of the drainage network must be derived from an understanding of the topography alone. Thus, one approach is to plot all the valley lines, whether or not a stream can be observed. This technique requires a well dissected terrain, and a relatively low sun-angle to accentuate the hill-side shading.

The contrast between lit and shaded areas is strong enough in unvegetated and snow-covered regions to allow a simple segmentation of the image into dark and light. Then, using a diagonal filter technique proposed by Seidel et al (1983), it is possible to accentuate the changes in brightness in the direction of the sun azimuth at the time of the image. The resulting image can be binarised into high values (light to dark i.e. ridges) and low values (dark to light i.e. valley lines). The sequence of operations is shown in Fig. 1.2 for the image of S. Wales in the snows of 1982. Wang et al (1983) use a similar valley-ridge model as a "front-end" for their spatial relational model discussed above.

Similarly, numerical techniques may be used to analyse digital terrain models to provide maps of valley lines, (Morris and Heerdegen, 1987) but it should be stressed that the pattern of valleys in an area may result from climatic patterns of geologic age, and may not represent current drainage characteristics at all well. One remedy is to digitise the drainage network, and add that information to the DTM in a geo-information system, (Moore, 1983). This has great benefits in the automation of streamflow estimation (Moore et al, 1987).

Thus, basic techniques have been proposed, by which the "segments" making up a drainage network may be defined and mapped on the basis either of lineaments assumed to be rivers, or on the basis of topography, assuming every valley to be a drainage line. The network itself may then be constructed, connecting where necessary on the basis of logical criteria. When all the networks in an area have been constructed then it will be possible to draw catchment boundaries between the fingertip drainage, and extract the quantiative features of the network needed in numerical models.

1.3 Methodology - assessment

Assuming that the techniques discussed above are successful at least to some extent, then a method of assessment has to be devised in order to evaluate them. The fundamental question here is whether the map of drainage network is a good one, and then, leading from that, there are the secondary questions of whether the numerical parameters extracted from the map are reliable in numerical models.

Firstly, then, we have to consider whether one set of lines (the interpretation) is a good representation of another set of lines (the map). Clearly, subjective criteria should not be used here, although it will be seen in the literature review that much of the earlier work is evaluated in that manner. The simplest objective criteria to employ is the "template" method i.e. if the interpretation and the map are brought to the same scale, and there are no more and no less lines on the interpretation than the map, and the interpretation can be rotated so that every part of every line occupies exactly the same position as the corresponding part on the map, then there is an exact correspondence, and the interpretation may be judged successful.

It is unlikely, though, that such a degree of excellence will be achieved, and a simplistic right/wrong assessment will not be appropriate. Having rescaled and rotated the interpretation to give the best fit possible to the template we may assess the goodness-of-fit in several ways. For example, we may measure the length of line in the map which has an exact correspondence in the interpretation, and calculate the statistic "percentage of map network correctly interpreted". However, it will be seen that even this simple assessment poses many difficulties. For instance, a small geometric distortion in the image could produce a network which is topologically similar to the map, but just displaced sufficiently to give a very low score (Fig. 1.3(a)). Whereas a minor error in the connections at the nodes of the system would give an interpretation which has a very high score, but which is radically incorrect (Fig. 1.3(b)).

It would be useful to retain some sense of the relative importance of parts of the network i.e. it is normally expected that the main branches of the network should be in the right place, whereas the exact location of the lower order streams is not so important. A useful, and simple measure would be that proportion of main streams which are located correctly to within some margin, say 3 or 4 meander widths : the definitions of "main stream" could be either the Water Authorities legal one, or it could be based on some measure of stream ordering.

This measure is robust, and a high score would indicate that the interpretation is basically sound. Unfortunately, a low score could still be misleading. For example, it is commonly the case that rail and road communications follow valley floors, crossing and re-crossing the rivers. Their reflectance properties are similar to rivers, and confusion between the two is easy. Thus it could be the case that the basic form of the network is mapped correctly, while the detailed location of the streams is consistently wrong (Fig. 1.3(c)).

Perhaps then, it is better to assess the topology of the main streams and ignore the details of location. An alternative way of measuring the "proportion of main stream length correctly mapped" is to examine the interpretation, and where a link between nodes is logically correct, then the length of that link is measured from the map. The nodes would be outlets, junctions and heads of streams.

Thus in Fig. 1.3(c), the interpretation would be said to give 100% correctly mapped main streams, whereas in Fig. 1.3(d) sections AB, BC and BD would be counted as correct, but all of EF would be incorrect. On a simplified example such as this the criterion appears harsh, but in real networks many more junctions would occur and the criterion would be tempered somewhat.

This measure has the merit of simplicity, and will form the basis for the method of evaluation proposed in the following sections. However, as noted above, it concerns itself solely with main streams, and should be used in conjunction with some measure such as stream frequency, which indicates the degree of success with lower order streams.

Clearly, stream frequency gives no indication of the topological correctness of the interpreted network, and if such an evaluation is needed then some other measure has to be devised. A method is now proposed in which stream ordering may be used to calculate a score, which indicates the degree of topological accuracy, but which is weighted, so that errors in fingertip drainage count less than errors in main rivers.

The procedure is to idealise the true drainage network as in the example in Fig.1. 4(a), and compute the Shreve stream order for each segment (Shreve, 1957). Then, idealise the interpreted network e.g. Fig. 1.4(c) and, working upstream, assign the 'true' order number to each new branch. Then calculate the sum of the Shreve order numbers for the whole network. In this example the true network has a score of 20 (Fig. 1.4(b)) and the interpreted network in Fig. 1.4(c) has a score of 17, while the rather poor interpretation in Fig. 1.4(d) has a score of 14. Expressed as a proportion of the 'true' score, this gives a correctness factor for (c) and (d) of 0.85 and 0.70 respectively. If only the main branch had been interpreted the score would have been 8 and the factor 0.40. Thus the method can be seen to give high scores for correctness of basic networks, and also give scores that are proportional to the number of low-order streams correctly plotted.

The main weakness with this method, as with any comparison with maps, is that any stream which has been detected on the image but not on the map scores zero, whereas it might actually exist, and might be shown on some map of different scale i.e. an improvement over the performance of the map is not credited. Clearly there are ways around this, but the method begins to be complicated.

An attractive method would be a combination of the last two, which would then bring together topological correctness and some evaluation of the length of network correctly mapped. The proposed technique for evaluating a "Correctness Factor" is as follows :

1. The interpreted network is drawn to the same scale as the true network.
2. The true network is idealised, and each segment is assigned its Shreve order number.
3. Working upstream, an overlay for the interpretation is created, on which stream segments are drawn between nodes, as long as the nodes on the interpretation are logically correct and are correctly positioned to within some margin. Streams from the interpretation are drawn between nodes if they follow the true

line to within some specified margin. The emphasis here is on "topological correctness"rather than exact line.

4. The length of each stream on the overlay is then measured, and is weighted according to the highest Shreve order calculated for that stream on the true network.
5. The same is done for the 'true' network, and a comparison made.

Thus main stream lengths will have a high weighting, reflecting their importance, while fingertip streams will have a weighting of unity. This method has the advantages that it provides a logical basis on which to evaluate the efficiency of an interpretation in providing maps of the drainage network in which increased emphasis is given to the more important parts of the drainage network. Its disadvantages are that it depends on the specification of an allowable margin of error, which may be an entirely subjective decision, and that it is complicated and time consuming to carry out. Its application is recommended where the shape and form of the drainage network is important, e.g. in making an inventory of surface water resources.

An alternative approach, which would be appropriate where the concern is for water resources modelling, is to set aside topological correctness, and evaluate those geomorphic parameters commonly used in hydrologic models. This is the approach most frequently found in the literature, and will be followed here, where the principal aim is to evaluate models. In setting aside topological considerations though, care is needed to ensure consistency of approach. For instance, in evaluating stream frequency, it must be the number of junctions in the catchment area as interpreted from the imagery, and not the catchment area as known from maps or water authority records.

Principally then, in the sections that follow, evaluations will be made of parameters most commonly encountered in models e.g. catchment area, main stream length, drainage density and also stream frequency, which is often used as a surrogate for drainage density because it is highly correlated, and is much quicker to measure. These will be evaluated firstly in terms of absolute values i.e. a straightforward comparison between the 'true' values estimated from maps, and the values estimated from the image interpretation. However, in most numerical models some calibration would be carried out, and so absolute accuracy is not as important as a good correlation with true values. Thus the evaluation will be mostly concerned with the correlation between true and interpreted values.

However, in situations where the accuracy of mapping needs to be considered, the evaluation will be made in terms of the Correctness Factor.

PART 2 CHAPTER 2 REVIEW OF LITERATURE

It is widely claimed in introductory material on remote sensing that satellite imagery distinguishes drainage systems with ease. Many full-scene images shown in such texts reinforce those claims. Yet very little has been done to confirm these impressions in any quantitative way, and certainly not in the more "difficult" types of terrain, such as those with low relief and dense vegetal cover.

The principal conclusion of the few trials, using Landsat MSS to delineate drainage basins and networks, is that it is successful only in areas of high relief. Rango et al (1975) attempted to measure basin area, basin shape and main channel sinuosity in a number of basins in three varied study areas of USA. They used ERTS-1 MSS imagery in single bands and also as colour composites, enlarged on a zoom transfer scope to 1:250,000 or 1:100,000 and compared the details to U.S. topographic maps at 1:250,000 scale. Multi-temporal data was used for one basin. Their findings were that catchment area could be found with acceptable accuracy (within 5-9%) in all their study catchments, but that delineation of the drainage network depended greatly on the relief of the area. In areas characterized by high relief and non-glaciated landscape, drainage density could be mapped in greater detail than that shown on 1:250,000 topographic maps. (In some cases, better than 1:62,500 scale maps). In glaciated areas, and where relief was moderate or low, they found that drainage density values were only 55% to 75% of those derived from maps. They noted that snow cover could greatly increase the ease of interpretation, and that the value of drainage density estimated from imagery could vary greatly with the date of the image.

Fowler et al (1977) made similar investigations, concluding that except in areas of very high relief, ancillary information from topographic maps is required. However, they were using optical equipment developed as part of a NASA project, which possibly compared unfavourably with a zoom transfer scope. They found seasonal observations particularly useful, and concluded that the Landsat imagery shows that effective drainage density varies cyclically throughout the year. However, in the light of the findings of Rango et al (1975), we might conclude that the variation in observed drainage density merely reflects a change in the ease of interpretation of the satellite imagery.

Far more encouraging results were achieved by Killpack and McCoy (1981) who compared geomorphic characteristics as measured from 1:24,000 topographic maps and 1:100,000 Landsat imagery. Using a zoom transfer scope they found that delineation of catchment features involved "a degree of subjectivity", but nevertheless claimed correlation coefficients of 97% for area, 98% for length of main channel and 96% for the length of drainage network in catchments of over 20 square miles. However, no indication was given as to the degree of relief in the topography, nor the density of vegetal cover in the catchments, which were all from one range of hills in Utah. They confirmed, though, the usefulness of winter imagery, in which topographical detail is enhanced by the shading produced by the lower sun angle.

In a European example, Astaras (1985) compared photoprints of Landsat MSS imagery at 1:250,000 scale with topographical maps of the

Hellenic Army Geographical Service at the same scale. He found a far greater number (897) of first order streams on the Landsat imagery than on the topographic maps (98), and a similar comparison for higher order streams. Stream lengths were also far greater when interpreted from the Landsat image (1132km compared with 757km for first-order streams). He confirms the benefits of sun-shading, and based his interpretation on the presence of open water, the presence of valleys or gulleys, or the presence of vegetation in gulley floors. His technique was to draw the drainage network onto a clear acetate overlay. Drainage network was extracted from the maps on the basis of blue lines, and on sharp crenellations in the contour lines.

With regard to other satellite systems, the resolution of the sensors is generally too coarse to be of value in catchment studies. An attempt was made by Schneider et al (1979) to map drainage features using the night-time thermal infrared imagery from NOAA-5, which was successful only in terms of the major features. However, it served to indicate the potential of the high resolution infrared sensors, such as those employed in the Landsat Thematic Mapper. No reports have been made as yet of the efficiency of the second generation satellite systems such as Landsat TM or SPOT, nor of the airborne thematic mappers such as the Daedalus 1268 scanner.

PART 2 CHAPTER 3

CASE STUDIES

This part of the study was concerned with assessing the value of visual interpretations of satellite imagery in making hydrologic maps of surface water. The literature review had already drawn attention to certain important issues, namely

(i) the date and quality of the imagery

(ii) the influence of topography and

(iii) problems of human perception.

The study was thus broken down in the following way :

Firstly a detailed study was made of the basic technique. This was based on 8 catchments in the South Wales area, and used Landsat MSS imagery. The basic study was extended in less detail to take in a variety of topography units in the South West of Britain to examine the effects of topography, and was repeated using imagery from four dates to examine the influence of image quality and seasonal effects. Certain interpretations were then repeated by other observers to evaluate the effects of subjectivity. This was followed by an examination of higher resolution imagery (simulated SPOT imagery) and finally a comparison was made with Landsat MSS imagery of an area with totally different geographic characteristics, namely Malawi in East-Central Africa.

3.1 THE MAIN STUDY- South Wales and the South-West

3.1.1 Description of method

This part of the study was based on four distinct landscape units adjacent to the Severn Estuary (Fig. 3.1), which all fell within Landsat scene 219/24. Within each unit an arbitrary rectangle was defined, and within one unit (South Wales Valleys) eight study catchments were defined. The principal landscape features of the units are given in Tables 3.2 and 3.3. The areas enjoy a similarity of climate (maritime westerly temperate), although rainfall is strongly linked to elevation. There is also a broad similarity in the main types of land use and so the principal difference between the areas is in their topography, which varies from the deeply incised parallel valleys of South Wales to the flat reclaimed wetlands of Somerset. The large variation in ruggedness of the topography is clearly indicated by the Horton slope number (Horton, 1932). A small area in North Wales (Afon Aled, Denbigh) was added, for which Landsat and SPOT was available.

The area used for the detailed part of the study was that part of South Wales lying in an around the South Wales coalfield. It is a well defined geographic unit containing a diversity of topographic forms, climate and land-use.

The study area is shown in Fig. 3.2 together with the 8 catchments for which data were available. The Usk and Honddu catchments lie principally on the Old Red Sandstone which dips gently to the South, giving rise to the steep escarpments of the Brecon Beacons and the Black Mountain, with a maximum height of 886m. The Neath and Taff rivers, including the Cynon, rise on the Old Red Sandstone, but pass over a major outcrop of limestone before crossing the South Wales Coalfield. The Coalfield is a large syncline in which the hard bands of Pennant Sandstone and Millstone grit stand out above incised valleys. The valleys are long, narrow and steep sided with a strongly parallel form. The southerly catchments of the Ogmore and Ewenny lie on the low plateau of the Vale of Glamorgan, with Liassic and Triassic rocks giving deep, fertile soils.

The climate of the area is of the Westerly Maritime type, with mild winters, cool summers and a persistence of rain throughout the year, but with a maximum in Autumn and Winter. Rainfall is closely linked with relief, varying from an annual average of 933mm in the coastal plain (Cardiff, Rhoose Airport) to 2400mm in the Brecon Beacons.

The high land is typified by thin, leached soils supporting a high proportion of grassland in the form of rough grazing. The valleys, which were once densely covered with deciduous woodland, were stripped during the industrial revolution, but are now being extensively re-forested with conifers.

The valleys of the coalfield are densely populated, but the settlements are wholly contained within the valley floors. Much of the area contains land reclaimed from the spoil heaps of the old iron works and coalmines. The reclaimed land has the thinnest of vegetation cover, and almost no soil. The Usk and Honddu basins are sparsely populated farming areas, while the Vale of Glamorgan has a higher population spread out in many small towns and villages.

For each unit the drainage network was mapped by visual interpretation of Landsat MSS imagery as photoprints of Band 7 (near infrared) at a nominal scale of 1:250,000. Cloud-free images were available for 4 dates, of which one was a winter scene, in which the whole area was covered by snow. The summer scenes suffered to varying degrees from the presence of haze and mist. The four images are shown in Fig. 3.3.

Maps were made from individual scenes, and also from a compilation of all available imagery. Drainage density (total length of channel per unit area) and stream frequency (number of junctions per unit area) were extracted for the whole of each unit. Catchment boundaries were mapped for the 8 basins of interest in the South Wales Valleys unit, then catchment area and stream frequency were evaluated for each basin. In the case of the Denbigh area, mapping was performed on just one Landsat scene at 1:250,000 scale, but a comparison was made with higher resolution imagery as reported in the next section.

The procedure employed in the mapping was to draw the drainage network onto a clear acetate overlay, using a separate overlay for each of the 4 images, and then using just one overlay for all of the images to build up a compilation. The process was repeated by each observer.

Features were distinguished as rivers if they possessed dark tone, were linear, and had typical features of rivers such as those listed by Wang et al (1983) as given in Table 1.1. The best sequence was found to be to draw in the obvious, large rivers, preferably working upstream from the coast line. These were generally large enough for the open water surface to give a distinctive dark tone, and were defined clearly enough for meanders and other typical river features to be visible. A positive identification was given according to the following criteria :

1. Dark tone
2. Linearity
3. Continuity
4. Bifurcations
5. Meanders

In the case of lower-order streams, they were not normally identifiable as a dark-toned linear feature. Thus a more intuitive approach was employed, taking into account the general features of the landscape. The following criteria were added :

6. Valley floor
7. Proximity of lake

The justification for using valleys to determine stream locations, is that this is a high rainfall area, in which streams are mainly persistent, except for a few small areas of karst topography.

A feature was given positive identification if it possessed all of features 1 to 5, and also "fitted" the topography. Streams were identified "with confidence" as extending from, or joining positively identified streams (i.e. by interpolation). They were also given that classification where a distinct valley was identified. (Very much detail of this type was extracted from the snow-cover scene, in which valleys were strongly accentuated by side-shading).

Streams were identified "with little confidence" as dark-toned irregular lineaments, possibly extending from positively identified streams, where there was little topographic evidence.

Reservoirs, lakes and ponds were identified on the basis of dark tone, and this was sufficient, even where topogrpahic context was of little help e.g. in the case of storage reservoirs.

Catchment boundaries were then drawn in, positioning them between fingertip drainage patterns, and using topographic information (e.g. obvious crest lines) wherever possible.

It was recognized that the technique adopted was very limited, and that in practice the hydrologist would have access to other information and techniques which would help in the interpretation. For instance, other documents such as aerial photographs may exist which show parts of the area at better resolution. Computer facilities may be available in which interactive enhancements of the imagery could provide a clearer rendition of the drainage network. In regression modelling, though, the key is consistency, and unless additional information is available at the same quality for the whole of the area it will not be useful in the regression.

3.1.2 Results - South Wales

The drainage networks and catchment boundaries of the South Wales area interpreted from a typical summer image, and the winter (snow-cover) image are shown in Figs. 3.4 and 3.5. The drainage network and catchment boundaries derived from the O.S. 1:250,000 scale map are shown in Fig. 3.6. Estimates of main stream length, catchment area and stream frequency taken from image interpretations, and O.S. Maps at 1:25,000, 1:50,000 and 1:250,000 scale are given in Tables 3.3 to 3.5.

Drainage network

In the summer scene (1980), the actual streams were discernible only if they were large, or if they were in areas of open moorland. The drainage network could be interpreted from topography in areas of high relief, but in areas of lower relief the interpretation depended largely on land-use. The areas of small-field farming and areas of forestry obscured much detail, even in areas of moderate relief. The winter image (1982) revealed considerable topographic details, but in a biased manner. Valleys in a southwest-northwest direction were delineated in far greater detail than those in the direction of the sun's rays.

The success of the interpretation in providing a reliable map of the drainage network may be judged by the Correctness Factor which was described above, and values of the Correctness Factor are given in Table 3.6, from which it can be seen that the winter image was quite reliable in providing locations of main drainage lines (although the fine detail in this interpretation did not contribute to the scores). The average score was 82% of the value from O.S. 1:250,000 scale maps. The summer image gave far poorer results, failing in some places to map the main drainage lines, and the average score fell to 54% of the map values.

Turning to estimates of geomorphic parameters, it can be seen that Table 3.3 shows that the 250,000 scale map and the Landsat interpretation both provide good estimates of Main Stream Lengths compared with the values given in the UK Flood Studies Report (NERC, 1975) which were extracted from 1:25,000 scale O.S. maps. Standard errors of estimate were 1.9km (map) and 2.8km (Landsat) for values ranging between 13 and 50kms. Even though there were significant errors in absolute terms, the correlation between Landsat and 25,000 scale was very good (r^2 = .89).

In making a comparison of stream frequency, however, the situation is rather more complicated. The comparison is again made with values quoted in the UK Flood Studies Report, which have been derived from 1:25,000 scale O.S. maps, and are given in Table 3.5. It is clear that the Landsat interpretations give absolute values which are far smaller than those extracted from 1:25,000 scale maps. However, they are of a similar magnitude to values estimated from 1:50,000 scale maps, and are greater than values extracted from 1:250,000 maps. It is clear that 'true' values of stream frequency depend on the scale of the map from which they are estimated. The important issue, then, is whether the values estimated from Landsat imagery give a good correlation with accepted values.

Figure 3.7 shows that there is no correlation at all between values of stream frequency extracted from the Landsat winter image and the Flood Studies Report. A possible explanation is that the amount of detail shown in the winter image depends greatly on the proportion of valleys which are given strong relief shading i.e. which lie southwest-northeast. If this proportion changed from catchment to catchment, then the estimated stream frequency also changed.

Surprisingly, the summer image, which shows far less detail, gives a much better correlation with Flood Studies Report. It is interesting to note that the performance of 1:50,000 and 1:250,000 scale maps as shown in Fig. 1.1 are worse than the summer imagery, with standard errors of estimate of 0.60 and 0.65 respectively. This may be a feature of this particular area, though, since Flood Studies Report indicates a much higher correlation in a national study (Figure 4.6 NERC, 1975).

Catchment area

The ability to sketch in the catchment boundaries depended almost entirely on the extent to which the drainage network has been defined. Where the catchment divide lay on the crest of an escarpment it was clearly shown in the winter image. Where this was not the case the imagery was of little help, although occasionally crest roads could be picked out. Frequently the imagery gave a strong impression of the underlying geology, which contrasted strongly with surface features.

A qualitative assessment can be gained from Figs. 3.5 and 3.8 which show that the summer imagery gave generally poor estimates of catchment boundaries. The winter image gave catchments of generally the correct shape (except the Ewenny catchment) but with numerous errors in the detail. The possession of summer and winter images did not help in the interpretation, but rather served to confuse matters.

A comparison of values of catchment area gives a simple quantitative assessment, but an important one, since that parameter is by far the most important in all types of hydrological models. The figures given in Table 3.4 indicate the accuracy of the imagery and the 1:250,000 map in estimating catchment area.

It can be seen from this that the use of quarter-million scale maps can give rise to significant errors in the estimation of catchment area, up to 6%. The use of Landsat imagery in this region gave results which were generally similar to those from quarter-million scale maps, but occasionally gave some extremely poor results, principally as a result of mis-identification of drainage networks. The correlation coefficient between Landsat 1:250,000 and O.S. 1:25,000 values of catchment area was 0.98, with a standard error of estimate of 20 km^2.

3.1.3 Results - South West

The drainage network identified for each of the five units using the clearest Landsat image (27 May 1977) is shown in Fig. 3.9.

Single date, single interpreter, effect of relief

The amount of drainage detail discernible in each unit is summarised in Table 3.7, which gives the drainage density for each area as calculated from the Landsat interpretation compared with values calculated from Ordnance Survey maps at 1:50,000 scale.

Firstly, it can be appreciated that drainage mapping gives values of drainage density of the same order as that from 1:50,000 scale maps. However, the efficiency of mapping is very variable, ranging between 45% and 96% of map values. Contrary to the impression of earlier researchers (Rango et al, 1975; Killpack and McCoy, 1981) no correlation is shown between drainage density and relief. Thinking that simple relief was too crude a measure, a comparison was made with Horton's Ruggedness number (Horton, 1932), but again there was no correlation.

Although no quantitative comparison can be made, it appears that efficiency of drainage mapping does relate to land-use. The highest values in Table 3.7 are for the areas containing much open moorland, whereas even in the steeply incised landscape of Denbigh the close texture of field patterns obscures the drainage network.

When making an evaluation of the topological accuracy of the mapping, it was noticed that in the agricultural areas such as the Usk-Wye vales there was a considerable confusion between streams and other cultural features such as roads, field boundaries etc., with the result that a high density of drainage was shown which was actually incorrect. The Correctness Factor for the sample square shown in Fig. 3.10 is actually as low as 26% of the value for the 1:50,000 scale map. By contrast, in the moorland regions the accuracy of mapping was much better, as can be seen in the sample square for Exmoor, which has a Correctness Factor of 75%. This highlights the danger of assuming that high values of drainage density indicate a good interpretation, as has often been assumed in the literature.

Effect of date; single interpreter, single region

Examining the interpreted values of drainage density for the 8 South Wales study catchments, no significant correlation was found with 1:50,000 scale map values on any of the 4 date images. This reinforces the point that interpretation of drainage networks from Landsat MSS is unreliable, and that high values of drainage density cannot be taken to imply good values.

Using catchment area as a more stable parameter, it can be seen from Table 3.8 that on average, estimates made from any single date image were reliable to within 6%, but that individual estimates were prone to quite large errors (standard errors of estimate up to $21.4km^2$). The best estimates were made for catchments in excess of $100km^2$, smaller catchment areas being estimated to be wrose than ± 10%. It was also noted that even where the area was estimated well, the shape factor could be up to 25% in error. The influence of date of image on estimate of catchment area appears to be very small, with correlation coefficients between true and estimated values being very close to 0.98 in all cases.

It should be remarked here that many of the catchments in this part of the study were adjacent to one another. A consequence of this is that an error in the location of a boundary would affect more than one estimate of area. Thus the standard error of estimate would be increased. Conversely, the estimate of mean values would be influenced very little.

Single date, single region, effect of subjectivity

Table 3.9 shows the stream frequencies estimated by the four observers, as a proportion of the 'true' value interpreted from 1:25,000 scale maps. In all cases the number of streams detected is far less than that on the map. This would be acceptable if the image values correlated well with map values. Unfortunately this is far from the case. Only one observer managed to achieve a positive correlation with map values, and even then the correlation coefficient was as low as 0.38.

Even though the detail of the networks was clearly incorrect, sufficient network was shown in all cases to attempt mapping of catchment boundaries. Individual success at boundary mapping appears to be independent of their success at network mapping and appears to be quite variable (Table 3.10). Correlations between true and estimated areas for each observer showed standard errors of estimate varying between 20 and $57 km^2$. Estimates of individual catchment areas by the four observers varied very widely, with one estimated catchment area having a range of 87% of its true value.

3.2 HIGH RESOLUTION IMAGERY

3.2.1 Description of method

This part of the study was aimed at investigating the benefit of higher resolution imagery, and was based on the Denbigh area, which was scanned as part of the National Remote Sensing Centre's 1984 SPOT-simulation campaign. Data were acquired with a Daedalus 1268 scanner, but processed to resemble that of the SPOT satellite i.e. 20m x 20m pixels in two visible bands and one in the near infrared.

The southern section of the Denbigh area lies in the Denbigh moors, which has broad, mature valleys in open moorland, but the nothern section is drained by steep-sided narrow valleys with arable land and woodlands in the valley floors.

The interpretation process was similar to that used in the Landsat study, except that photoproducts were made at 1:50,000 scale, and that the networks were compiled from the near infrared image and the 3-band false colour composite. The images resembled aerial photography and were sufficiently clear for cultural features to be easily recognized, so that there was no need to establish a set of criteria for defining rivers.

3.2.2 Results

Interpretations were made for the Afon Aled catchment, which lies in the northern part of the area i.e. it is a deeply incised network. The drainage networks interpreted from SPOT-simulated imagery and Ordnance Survey maps at 1:50,000 scale are shown in Fig. 3.11.

The main drainage lines were mapped correctly, but there was a considerable amount of error in the lower-order streams, where there was confusion between streams, roads and narrow patches of woodland (Fig. 3.12). The Correctness Factor was 82% of the 1:50,000 scale map value, which compares well with the Landsat interpretation for South Wales at 1:250,000 scale.

Drainage density estimated from the image was 71% of the 1:50,000 map value, which appears to be low. However, it was felt that this was a particularly difficult area to interpret. Making a comparison of interpretations at 1:250,000 scale using Landsat MSS (Table 3.11) confirmed this impression.

The conclusion, then, is that simulated SPOT imagery was interpreted to provide maps and drainage density estimates that are somewhat poorer than 1:50,000 maps, but that in other areas SPOT Imagery can be expected to provide much better details.

3.3 A SEMI-TROPICAL DRAINAGE BASIN

It was recognized that the study areas reported in 3.1 and 3.2 all lay within the same climatic and hydrologic regions. Thus the study was extended to examine whether the technique gave similar results in a drainage regime which was radically different.

The study area that was chosen was that part of the central plateau of Malawi lying between 35° and 34° East and 12° to 13° South. It is an ancient erosion surface showing very subdued relief, on which ferruginous soils support a natural vegetation of Brachystegia-Julbernadia woodland (i.e. "bush"). Rainfall is typically 1200mm per annum, occurring principally from late November to mid-March.

The area is drained by two major rivers, the Dwanga and Bua, whose flows are measured at well-established gauging stations. The catchment of the Dwanga upstream of the gauging station lies almost entirely in a National Game Park, and is covered in undisturbed natural vegetation. The catchment of the Bua, however, lies in an area which is extensively cultivated both for subsistence agriculture (maize gardens) and for cash crops (tobacco estates) and thus provides a striking contrast in hydrology.

Hydrological data were available for the Dwanga and Bua rivers and for the Rusa river which is a major tributary of the Bua. (Drayton et al, 1980). The Rusa river has a very similar catchment area to the Dwanga and was thus used for comparison.

Good quality topographic maps were available at 50,000 and 250,000 scale (Malawi Government Dept. of Surveys). The only satellite imagery available was a Landsat MSS image, scene 181/69, taken on 18th November 1981. This date was fortuitous since it tied in well with the latest 50,000 scale map edition (1981), and was at the best time of year for identifying dambos i.e. at the end of the dry season.

Dambos are a very important feature of the drainage of this area. They may be defined as "the periodically inundated grass-covered depressions on the head water end of a drainage system in a region of dry forest or bush vegetation", (Ackerman, 1936). The importance of dambos lies in their ability to act as natural reservoirs, storing large quantities of water during the rainy season, and releasing it slowly through the long dry season. In doing so they provide sites for the cultivation of essential dry-season dietary supplements such as green vegetables. They also provide grazing for cattle, and all too often form the only source of potable water, usually extracted from primitive shallow wells.

To the hydrologist their importance is in the way dambos condition the response of the catchment to rainfall, suppressing the flood flows, and enhancing the reliability of river flows further downstream. The proportion of catchment area which is dambo also determines to a large extent the magnitude of the losses from evapotranspiration and thus determines the yield of the catchment. The importance of dambos in regional development and in water resources development is thus clear, and a number of basic studies have recently been reported e.g. Mackel, 1985; Gregory, 1986; Faulkner, 1986.

Fundamental to dambo research is the need to be able to map them. In areas of natural vegetation dambos are very distinct broad avenues of grassland bordered by woodland on the interfluves. Their length and width is thus easily defined. In areas which have been taken over for cultivation very little tree cover may be left, and the location and size of dambos is often very hard to determine.

Thus in this part of the study the aim was to assess the performance of the technique not only in estimating the length of the drainage system, but also in estimating the width, and hence, area of dambo in the catchment.

3.3.1 Method of study

Photoprints at a nominal scale of 1:250,000 were prepared, one showing Band 7 (near infrared) only, and one showing a conventional false colour composite (Bands 4,5,7: blue, green, red). Drainage lines and dambo boundaries were then traced onto a clear acetate overlay. In practice the false colour composite proved much easier to interpret and the Band 7 image was not used.

Identification of the drainage features proceeded as follows. Where the rivers were wide enough, and clear of vegetation, they could be seen as lineaments having the sinuosity typical of a meandering river, and exhibiting the dark tone typical of an open water surface. In some cases the river channels were wide enough to be resolved by the satellite sensors, but were covered with vigorously growing vegetation (e.g. papyrus) and thus showed as bright red lineaments. Dambos showed as broad lineaments having a dendritic arrangement but without the sinusoity of a river channel.

Identification of dambos by tone depended on several factors. Dambos by definition are treeless areas and in the forested Dwanga catchment they appeared as areas of dry weaker vegetation (grass) bordering the dark-toned moist areas remaining at this time in the dry season. In the cultivated Bua catchment the interfluves were generally bare soils prepared for planting, and were dry i.e. showing a blueish range of colours in false colour. By contrast the dambos were still moist (dark tones) or still supported grass growth (red colour) and were thus easily identified. Problems arose in areas where there had been bush fires, which showed in very dark tones, sometimes obscuring the dambo network.

Having identified drainage lines and dambo areas, they were drawn onto the clear overlay and numerical parameters such as catchment area, drainage density and dambo area were evaluated.

RESULTS

Figure 3.13 is a photograph of the false colour composite showing the junction between the forested area (to the north) and the cultivated area. The very extensive Rusa dambo is clearly evident. Figures 3.14 and 3.15 show the drainage networks of the Rusa and Dwanga rivers as interpreted from the false colour composite at 1:25,000 scale, and as copied from the 1:250,000 scale maps.

It appears from Figs. 3.14 and 3.15 that the Landsat interpretation has yielded a similar amount of detail to the 1:250,000 scale map in the Rusa catchment, but has shown considerably more detail in the Dwanga catchment. Checking the topological accuracy of the interpretations it was found that in both cases the drainage network as shown on the maps is almost exactly delineated on the image interpretation. Correctness Factors of 99.8% and 100% were found for Rusa and Dwanga respectively.

The accuracy of the mapping may be partly attributed to the ease in which the very broad dambos could be differentiated from the surrounding vegetation. Another important factor, though, is the relative absence of other cultural features such as roads, railways, canals and hedgerows which cause such great confusion when interpreting imagery of a highly developed area such as South Wales. This reinforces the conclusion made previously that success in mapping is related to land-use, and suggests that maps made in areas of uniform 'natural' vegetation are likely to be more accurate than those made in developed areas.

The values shown in Table 3.12 confirm that in the Dwanga catchment very much more detail was shown than in the 1:250,000 scale map. Considering the uniformity of geology, climate and soils throughout the region it was thought that an actual difference in drainage density between Dwanga and Rusa was unlikely. It was not possible to acquire field data, and so the only alternative was to check the 1:50,000 scale maps of the region even though they were known to be somewhat unreliable. Samples of 10km grid sqares were taken from both catchments, and a typical square is shown in Fig. 3.16.

The detailed mapping confirmed the satellite interpretation, showing that drainage within the Dwanga basin was very much more dense than indicated on 1:250,000 scale maps. Drainage in the Rusa catchment was shown to be confined to a small number of drainage lines, but with a very much greater width of dambo. The Landsat interpretation was good enough to show a length and area of dambo slightly in excess of that on the 1:50,000 scale maps (Table 3.13).

PART 2 CHAPTER 4 SUMMARY AND CONCLUSIONS

The first impression on looking at the imagery was that drainage networks stand out clearly and that drainage mapping should be easy. Unfortunately, it is topography, not drainage that stands out, and the correlation between the two is not exact.

Attempts have been made to devise strategies for the automatic interpretation of lineaments by machine, but the computer-vision problems are hard to solve, and progress is slow. In this study the interpretation was made visually using photo products of imagery at 1:250,000 scale, and results were compared with maps at 1:250,000, 1:50,000 and 1:25,000 scale. Objective criteria were established for the definition of streams and rivers, that were based on criteria that had originally been developed for machine interpretation. It was found that it was essential to use objective criteria when interpreting Landsat imagery at 1:250,000 scale, and that they needed to be revised for different hydrologic regions. Interpretation of high-resolution (e.g. SPOT) imagery could be done using normal air-photo interpretation techniques, and objective criteria were not needed.

Drainage maps were prepared for eight basins in the South Wales region, where it was found that ease of interpretation depended to some extent on the relief of the area, but also to a large extent on the land-use of the area. A method was proposed for assessing the quality of drainage maps interpreted from imagery, but the maps were also assessed on the basis of values of geomorphic parameters estimated from them. Examining one image for 8 basins, the following average values and correlation coefficients were found :

Landsat 1;250,000 interpretation	1:250,000 map value	1:25,000 map value		
		%	r^2	s.e.e.
Main stream length	93%	80	96	2.8
Catchment area	93%	91	98	20.0
Streams frequency	222%	6	52	0.03
Correctness Factor	82%			

It was noted that the interpretations showed many topological errors. This was confirmed by the Correctness Factor, which showed that many streams were mapped, but that they often bore little relation to the true network. Nevertheless, errors in the shape of catchment boundaries tended to even out, so that the estimated catchment areas correlated well with true values as did main stream length.

The accuracy of stream frequency (or drainage density) varied strongly from catchment to catchment, and was seen to vary strongly between different topographic units, but with little obvious connection with relief. There was no strong correlation with true values in any land unit, nor for any of 4 different observers. However, some correlation was found using poor quality images, where only the major rivers could be discerned.

The effects of subjectivity was most noticeable in stream frequency, where large differences in absolute values and correlation were noted between four observers. It appeared, though, that errors in the details of the network affected the shape of catchments, giving errors in shape factor of up to 27%, but had little affect on the estimate of catchment area, which was consistently estimated well by all observers on all images.

Given the success and reliability with catchment area and main stream length, it is concluded that satellite image interpretation is a useful tool for regional hydrologic studies, and simple modelling, where area is by far the most important parameter. Success was achieved in simple tasks, but the technique failed to provide detailed information reliably.

Looking at a limited amount of higher resolution imagery, it became apparent that the method of interpretation was radically different, based on recognisable cultural features. In an area that was confirmed as being difficult to interpret,a drainage density of 71% of the true value was achieved, comparing at 50,000 scale. Clearly, more work will be needed on SPOT imagery when it is available, but it looks hopeful for mapping at sub-regional level.

The results given above correspond to those given in the literature for North American catchments. In an attempt to examine the technique in a sub-tropical region in Africa, stream frequency averaged 174% of that shown on 1:250,000 scale maps (compared with 222% for the South Wales region). However the topological accuracy was significantly better, partly due to the very wide zone which is associated with the drainage lines in that region, and partly due to the absence of the many cultural features which causes confusion in the South Wales scene. An important parameter, area of dambos, was found to be estimated well at 1:250,000 scale.

Thus it appears justifiable to extend the conclusion that the technique is suitable for regional studies to include other parts of the world.

Conclusions

The visual interpretation of Landsat Band 7 images at 250,000 scale is a quick and cheap method of preparing maps of the main river networks of a region, and extracting reliable values of catchment areas. The techniques could be easily learnt by anyone with experience in air-photo interpretation and requires no special equipment. It is thus highly suitable for preliminary water resource studies in the Third World, and is appropriate to simple, regional hydrologic models.

Catchment area is the most important parameter in regional hydrologic models, and estimates of area are remarkably insensitive to either the quality of the image or the nature of the terrain. Estimates of area made from Landsat Band 7 images appear to be so reliable that there is little benefit to be gained in using either enhanced Landsat composites or SPOT imagery, either of which would be more expensive.

The mapping of low order streams using single date, single band images is inconsistent, and gives rise to a poor correlation for numerical parameters of drainage density when compared with map values. The reliability of estimates is strongly related to date of image, relief shading and land cover. Very significant improvements in reliability can be achieved by compiling details from several different date images, which gives very little increase in cost.

In practice, the detail of drainage maps could be improved using other sources of information such as photographs, or by interactively enhancing the imagery. It is unlikely though that improvements would be gained in regression models, where consistency of data quality is paramount.

Unit	Max/Min Elevation (m)	Drainage Structure	Horton Slope No.	Area (km^2)
S.Wales Valleys	850/0	Parallel	0.107	1600
Usk-Wye Vales	530/15	Rectangular	0.062	1200
Exmoor	480/120	Rectangular	0.093	1600
Somerset Levels	290/0	Largely artificial	0.017	1200
Denbigh	520/105	Dendritic	0.146	123

TABLE 3.1 Physiographic details of study area

Unit	Land-use
S.Wales Valleys	High moorland, forest and industry in valleys
Usk-Wye Vales	Mixed agriculture, small fields
Exmoor	High moorland, lowland and farming
Somerset Levels	Reclaimed wetlands, meadow
Denbigh	Mixed agriculture, small fields

TABLE 3.2 Land-use within study areas

Catchment	MSL (km) (FSR)	MSL (km) (250,000 map)	MSL (km) (Landsat)
Honddu	20.2	17.5	17
Lwyd	26.8	22	20
Usk	22.5	17	21
Cynon	25.8	24	17
Taff	42.3	36	35
Ogmore	20.2	19	16.5
Neath	28.4	26	22
Ewenny	13.6	13	-

TABLE 3.3 Comparison of Main Stream Length derived from UK Flood Studies Report, O.S. 1:250,000 map and Landsat imagery

Station	FSR	1:250,000 Map	Summer Landsat	Winter Landsat
Honddu	62.2	60.2	18.8	63.5
Afon Lwyd	98.1	91.8	120.6	97.3
Usk	184.0	186.3	278.1	200.0
Cynon	109.0	104.4	56.3	98.5
Taff	455.0	444.8	398.8	458.1
Ogmore	158.0	154.5	146.9	137.9
Neath	191.0	188.1	192.5	137.5
Ewenny	62.9	60.6	-	-

TABLE 3.4 Estimates of catchment area made from UK Flood Studies Report, O.S. 1:250,000 map and Landsat imagery)km^2)

Station	FSR	1:50,000 Map	1:250,000 Map	Summer Landsat	Winter Landsat
Honddu	1.01	0.47	0.080	0.05	0.457
Afon Lwyd	1.17	0.77	.020	0.083	0.329
Usk	1.67	0.22	.108	0.104	0.330
Cynon	2.33	0.72	.037	0.178	0.264
Taff	2.17	0.78	.068	0.138	0.220
Ogmore	2.63	0.99	.082	0.204	0.370
Neath	2.59	0.87	.136	0.062	0.109
Ewenny	1.41	0.83	.079	-	-

TABLE 3.5 Estimates of stream frequency made from UK Flood Studies Report, O.S. 1:50,000 and 1:250,000 maps and from Landsat imagery (junctions/km^2)

Catchment	Correctness 1982	Factor 1980
Honddu	98	30
Llwyd	74	69
Usk	90	38
Taff (including Cynon)	88	87
Ogmore	No readings	
Neath	80	72
Ewenny	59	29

TABLE 3.6 Values of the mapping 'correctness factor' given as a percentage of values from 1:250,000 scale maps

Unit	Drainage Density (%)	Relief (m)	Horton Ruggedness
S.Wales Valleys	73	850	0.11
Usk-Wye Vales	68	515	0.06
Exmoor	96	360	0.09
Somerset Levels	62	290	0.02
Denbigh	45	415	0.15

TABLE 3.7 Drainage density from May 1977 image as percentage of value extracted from 1:50,000 O.S. Maps (single interpreter, single date)

	Aug. 80	July 77	May 77	Jan 82	Compilation
Mean	0.94	1.01	0.99	0.97	1.00
S.D.	0.16	0.17	0.13	0.17	0.20
r^2	0.98	0.98	0.98	0.99	0.98
s.e.e.(km^2)	20.0	19.8	21.4	15.8	20.8

TABLE 3.8 Estimated catchment area as proportion of true catchment area. (8 South Wales catchments, single interpreter)

Observer	1	2	3	4
Mean	0.11	0.06	0.05	0.13
S.D.	0.05	0.02	0.03	0.08
r	-0.10	0.62	-0.18	-0.98
r^2 (km/km)	0.01	0.38	0.03	0.95

TABLE 3.9 Estimates of stream frequency as proportion of true value (8 South Wales catchments, single date)

Observer	1	2	3	4
Mean	0.94	0.91	0.94	0.99
S.D.	0.14	0.37	0.35	0.17
r^2	0.98	0.84	0.91	0.97
s.e.e. (km^2)	20.0	56.9	40.4	22.0

TABLE 3.10 Estimates of catchment area as proportion of true value (8 South Wales catchments, single date)

	Drainage density from 250:000 Landsat MSS c.f. 250,000 O.S. maps
Aled	94%
Rheidol	48%
Usk	193%
Exmoor	193%

TABLE 3.11 Comparison of ease of interpretation of different areas

	Rusa		Dwangwa	
	Landsat	Map	Landsat	Map
Length of main river (km)	1048	916	2150	1523
Catchment area (km^2)	2380	2580	2914	2980
Drainage density(km/km^2)	0.406	0.355	0.721	0.304
Stream frequency(j/km^2)	0.094	0.078	0.290	0.127

TABLE 3.12 Comparison between Landsat MSS interpretation and Malawi Department of Survey maps at 1:250,000 scale

Parameters	Value expressed as % of value obtained from 1:50,000 map
Drainage density	110
Stream frequency	104
Length of dambo	114
Area of dambo	110

TABLE 3.13 Comparison of estimates from Landsat interpretation and 1:50,000 scale map (Average of 4 10km x 10km squares from Dwanga and Rusa catchments)

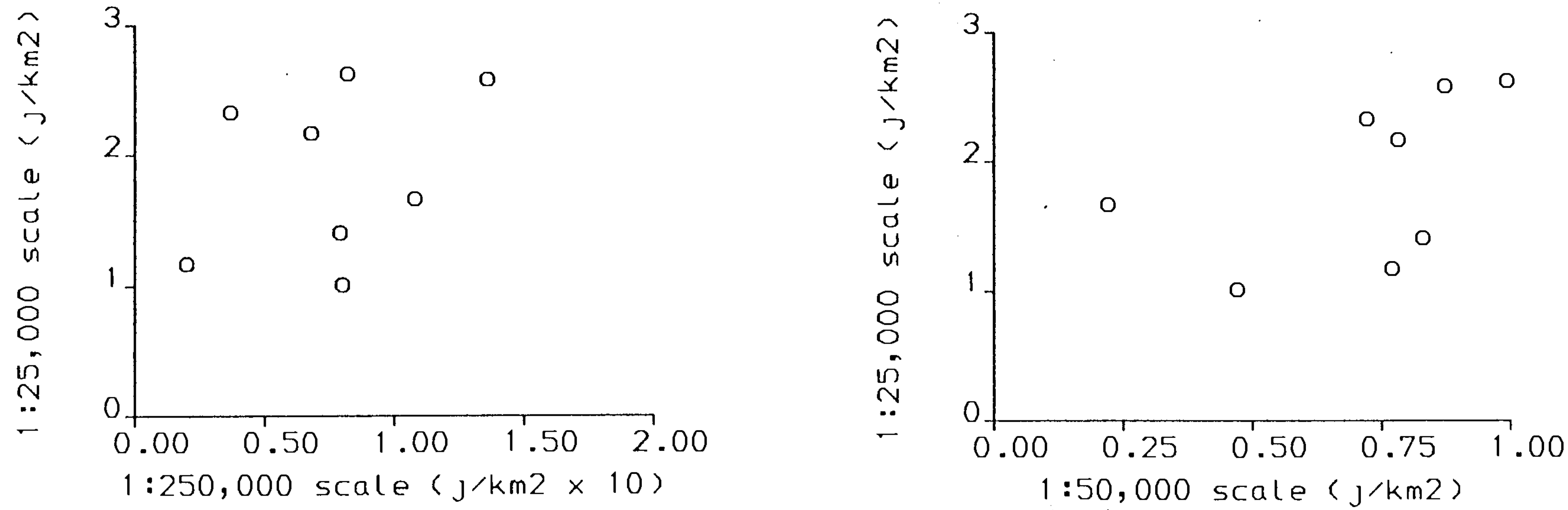
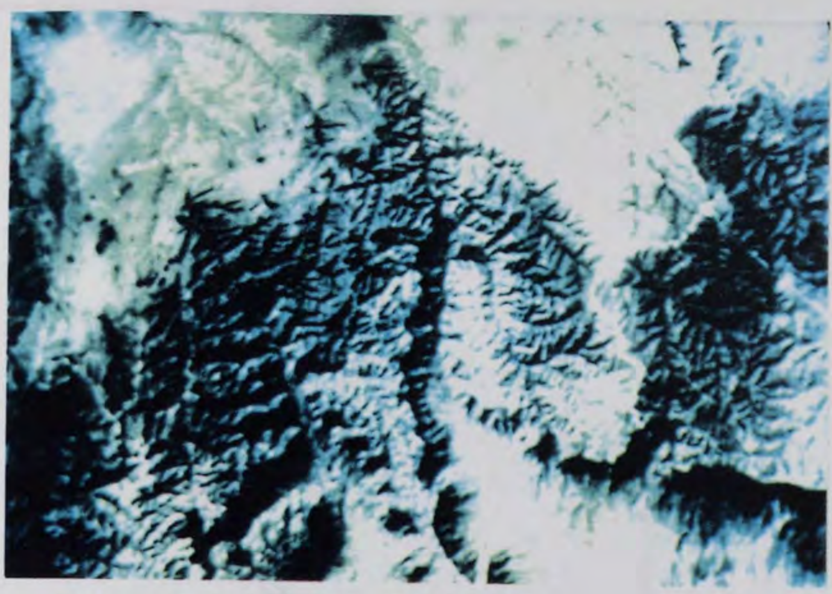
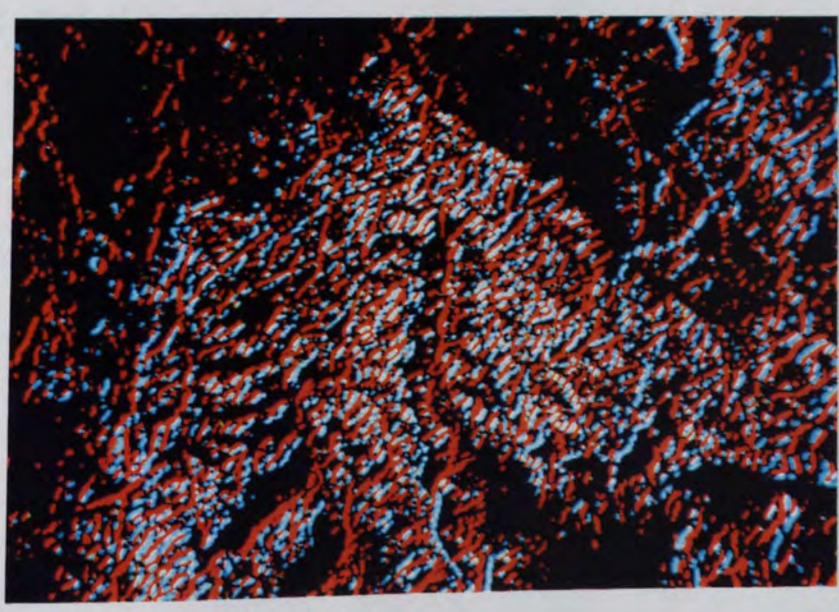


FIG. 1.1 COMPARISON OF STMFRQ VALUES FROM MAPS



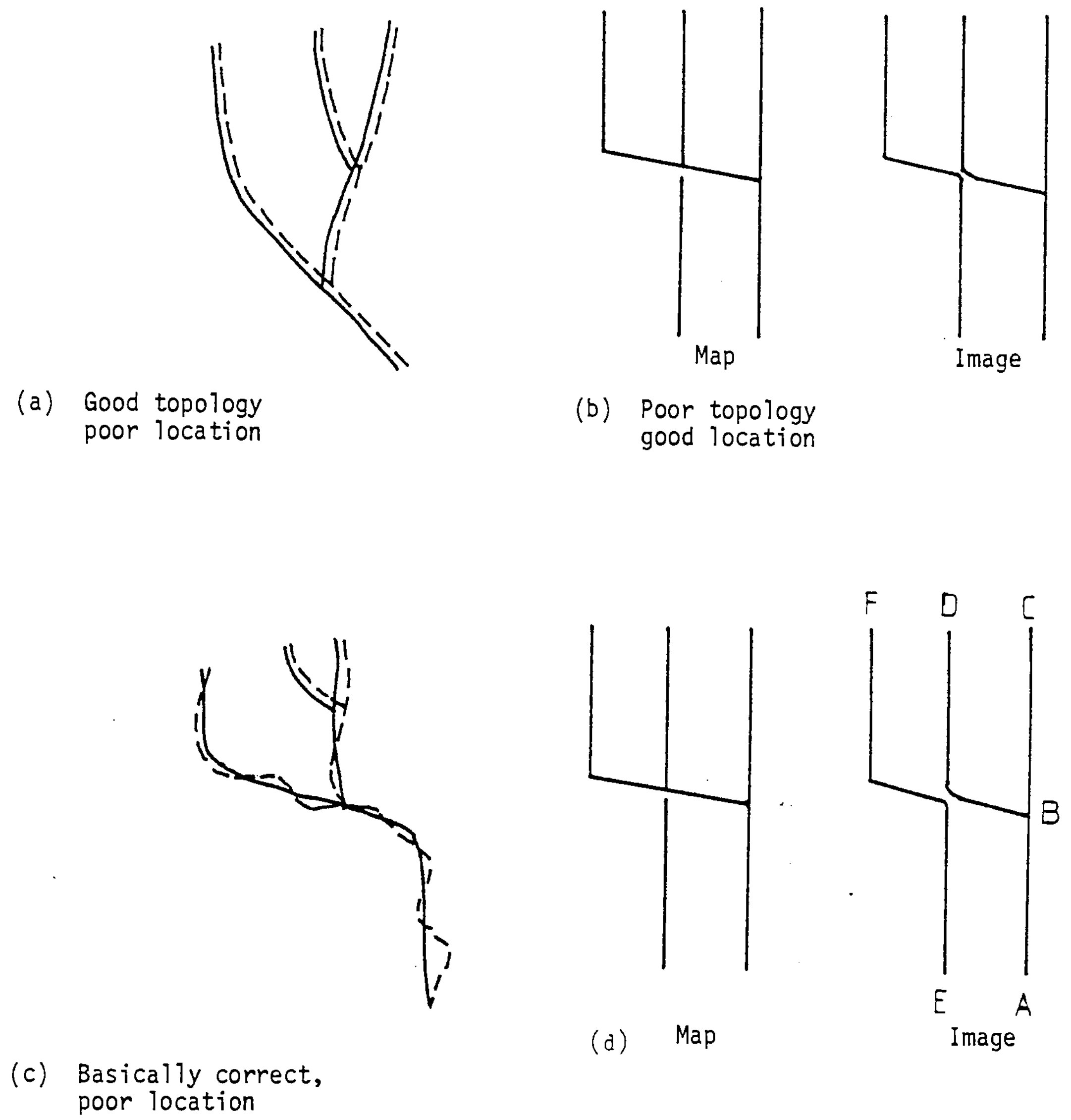
(a) Band 7

(b) Diagonal filter



(c) Density sliced

Fig. 1.2 Extraction of valley and crest lines



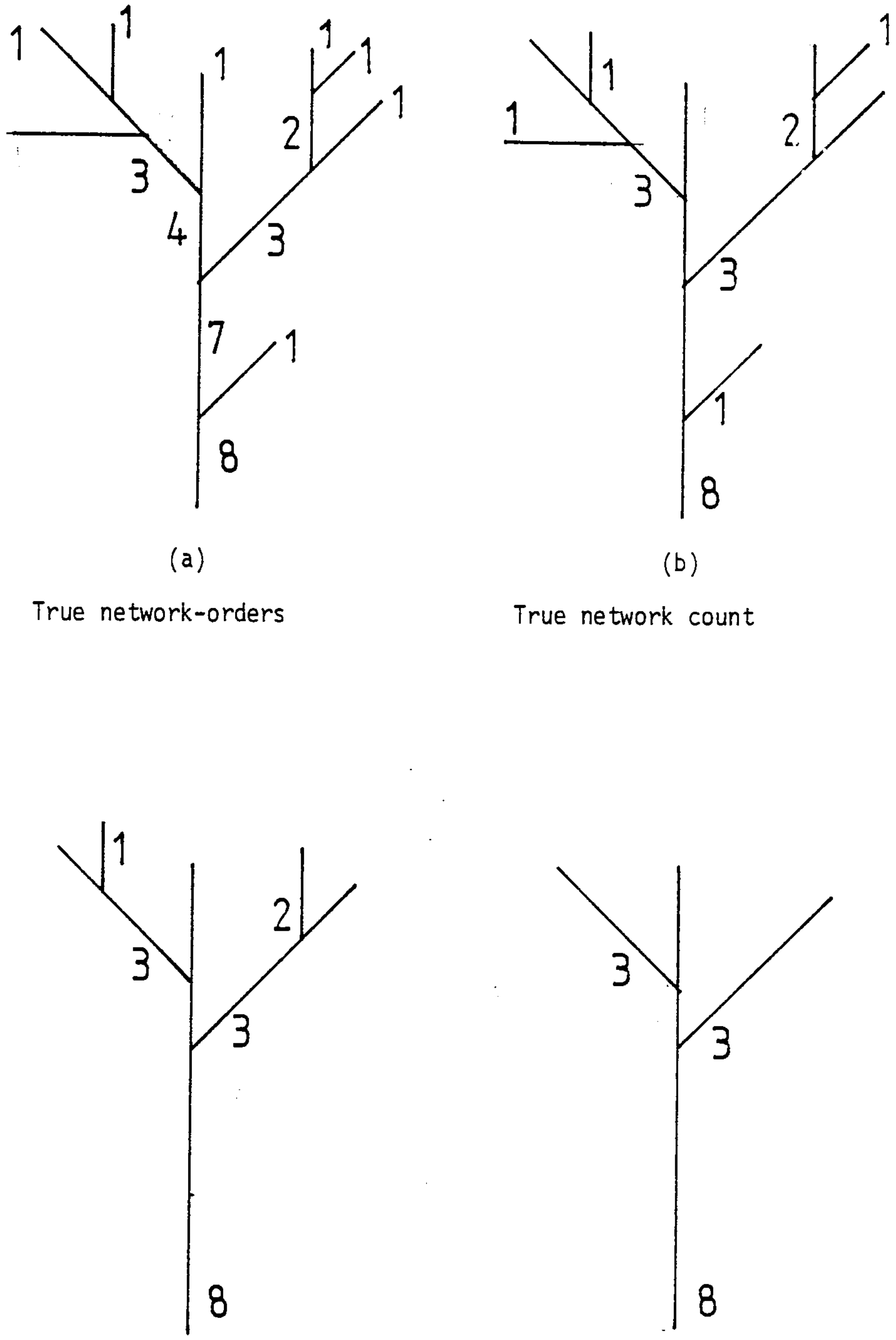
(a) Good topology poor location

(b) Poor topology good location

(c) Basically correct, poor location

(d)

Fig. 1.3 Correspondence of mapped networks with interpolation



(a) True network-orders

(b) True network count

(c) Interpreted network count (good)

(d) Interpreted network count (poor)

Fig. 1.4 Method for calculating the 'correctness score'

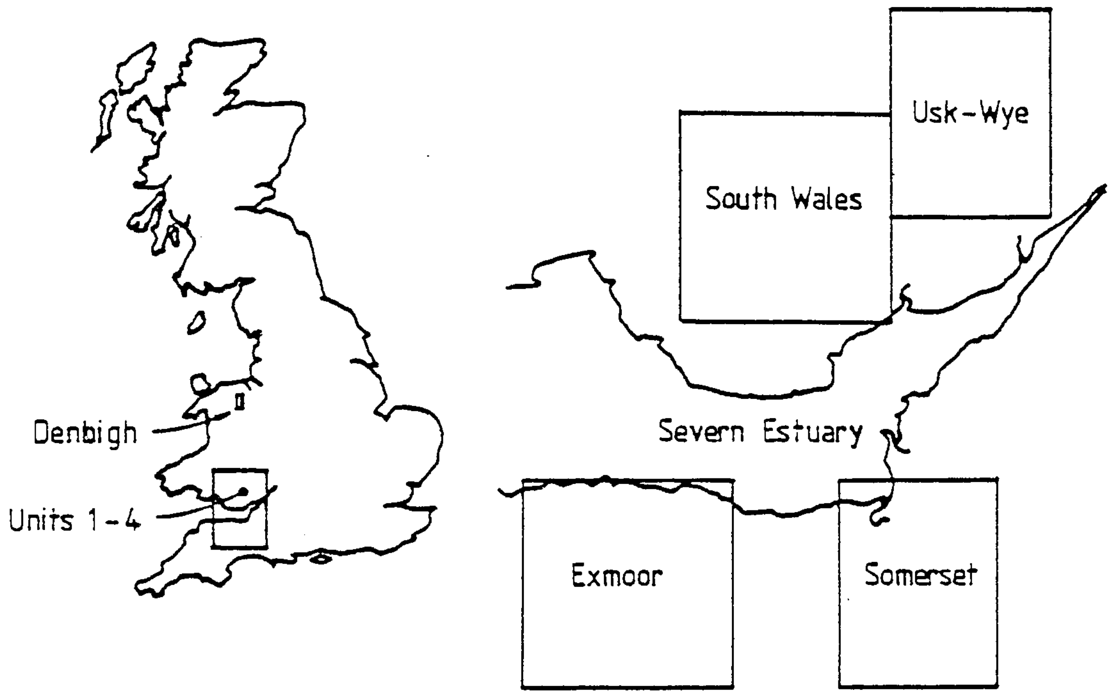


Fig.31 Location of study areas

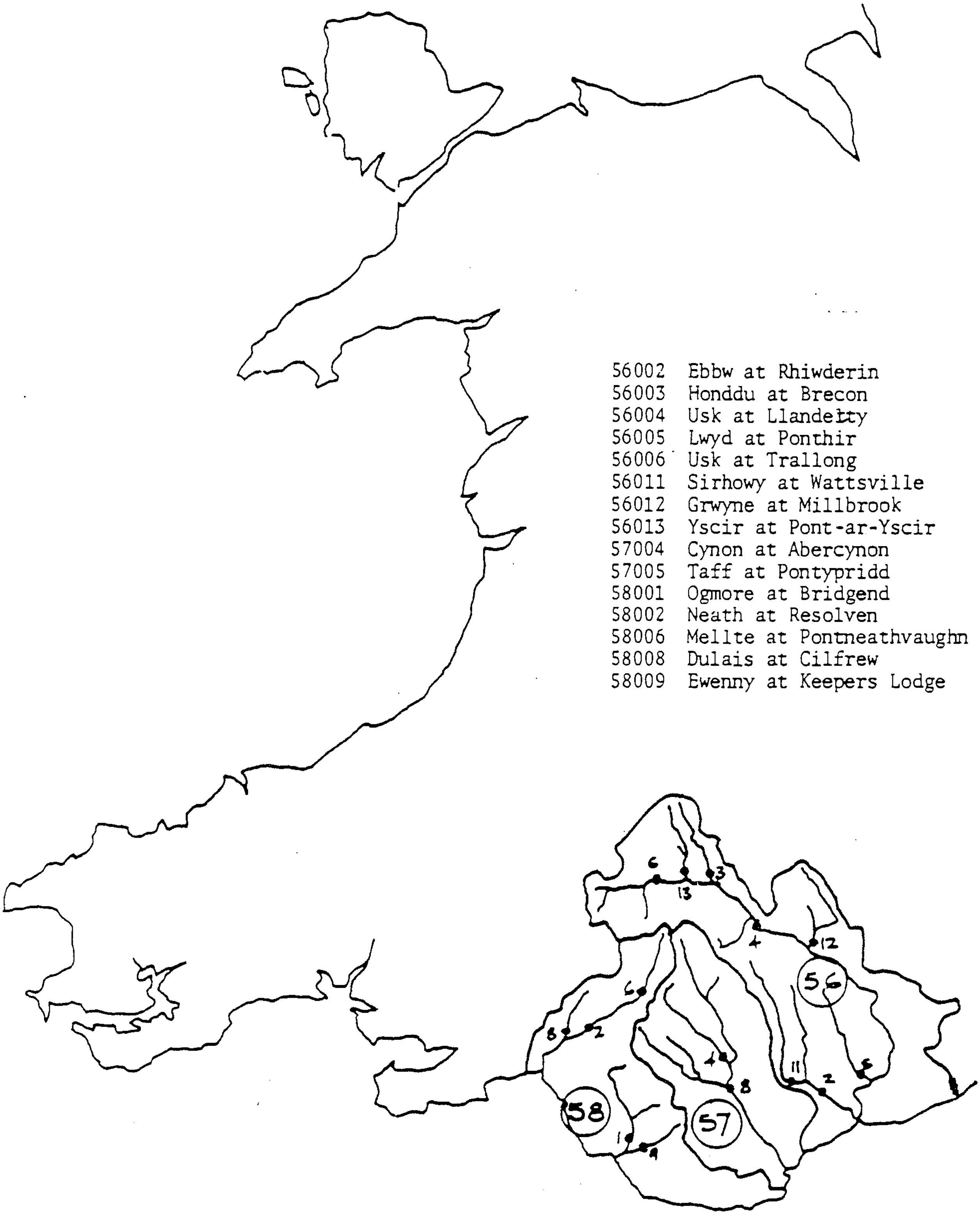
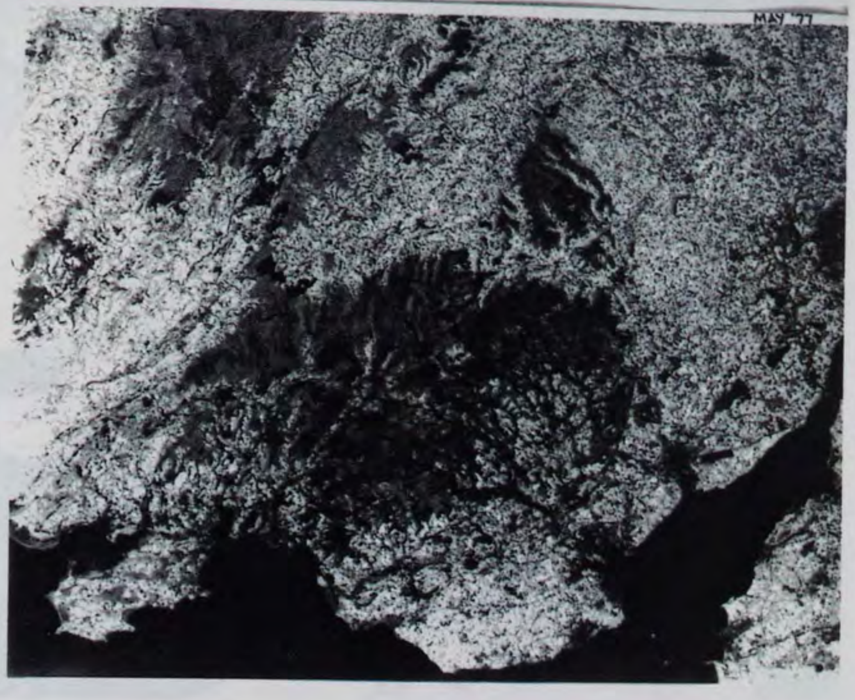
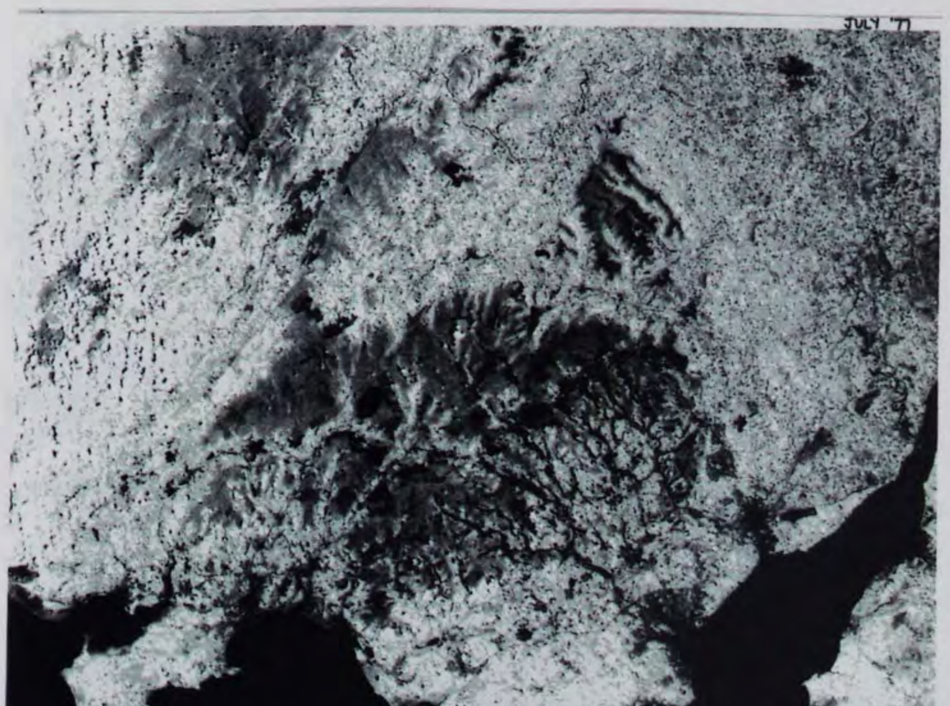


Fig. 3.2 STUDY CATCHMENTS

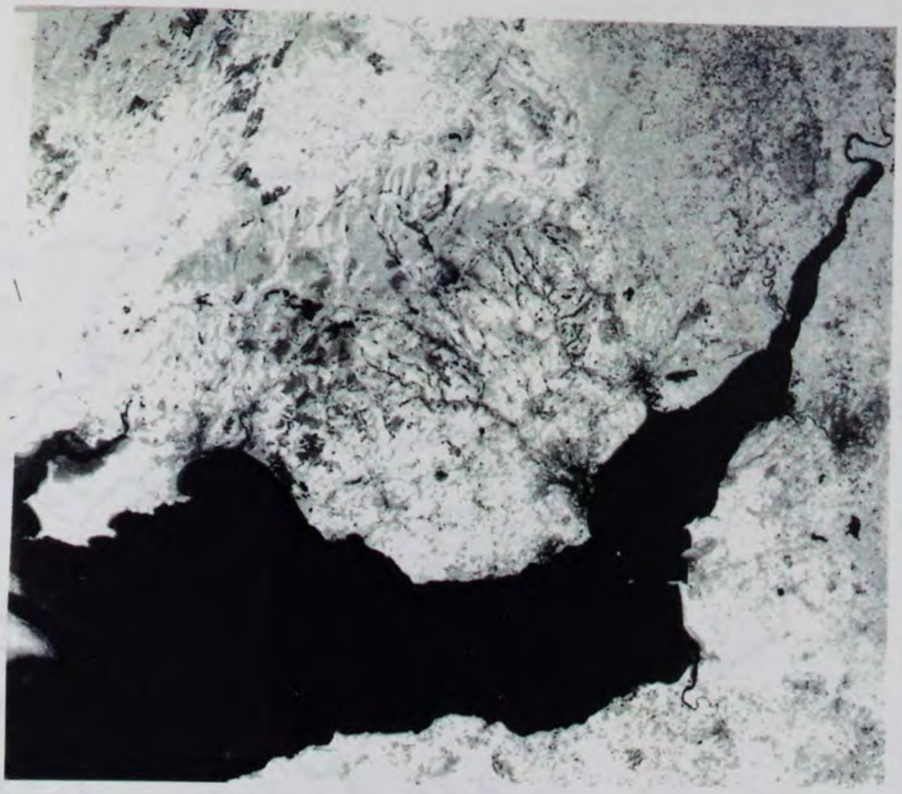


(a) May 1977

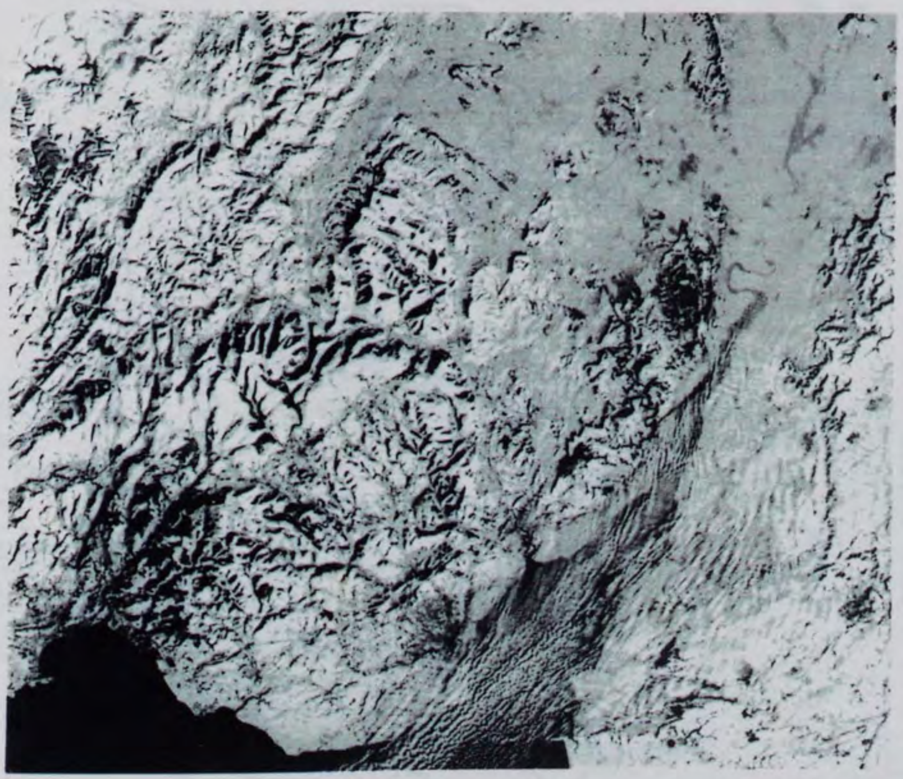


(b) July 1977

Fig. 3.3 South Wales Imagery

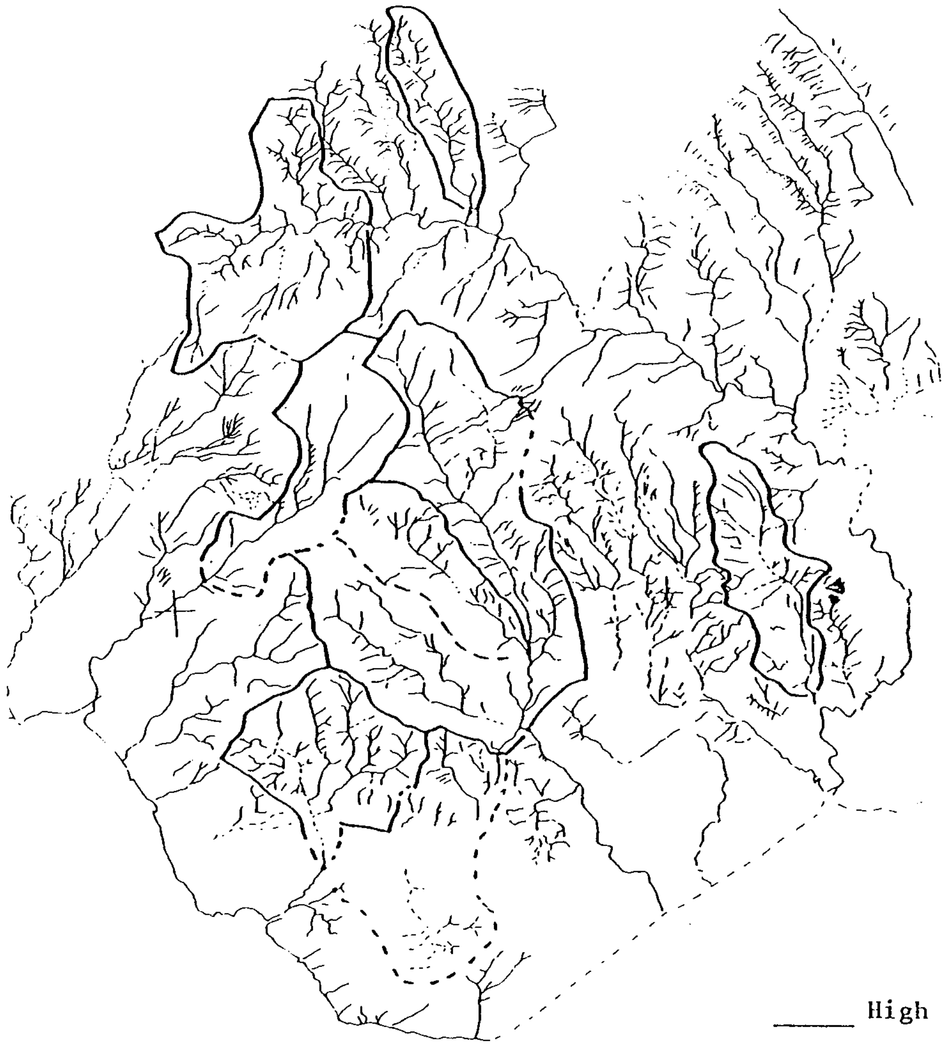


(c) August 1980



(d) January 1982

Fig. 3.3 South Wales Imagery



_____ High confidence

----- Low confidence

Fig. 3.4 Winter 1982

Fig. 3.5 Summer 1980

Figs. 3.4 and 3.5 Drainage network and catchment boundaries interpreted from Landsat Band 7

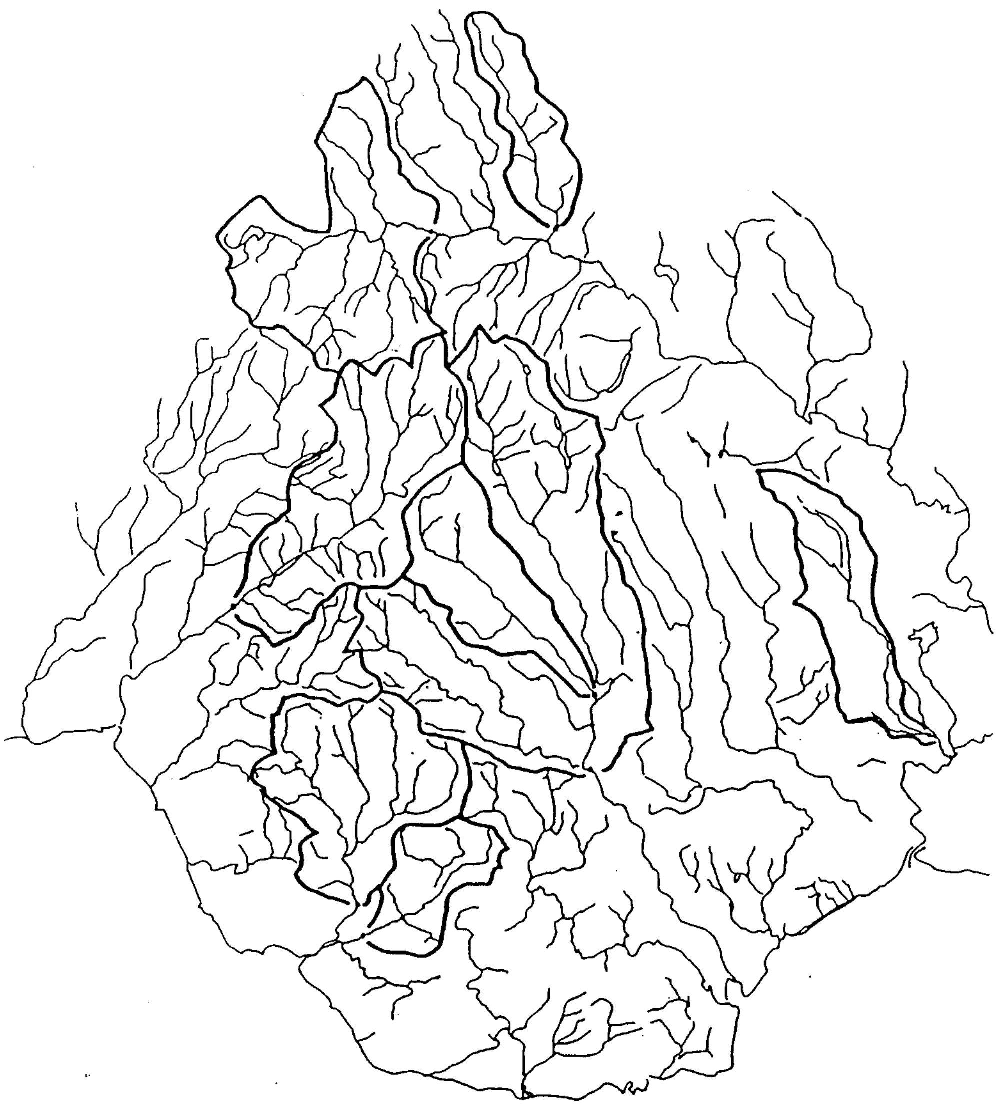


Fig. 3.6 Drainage network and catchment boundaries interpreted from 1:250,000 map

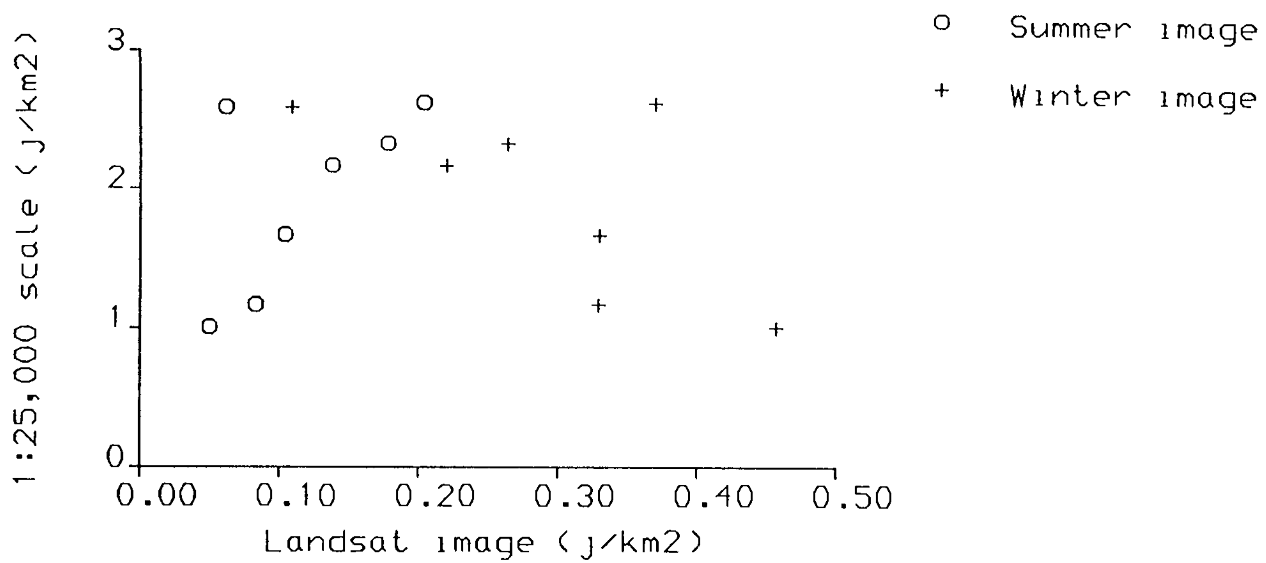


FIG.3.7 COMPARISON OF STMFRQ VALUES FROM IMAGERY AND MAPS

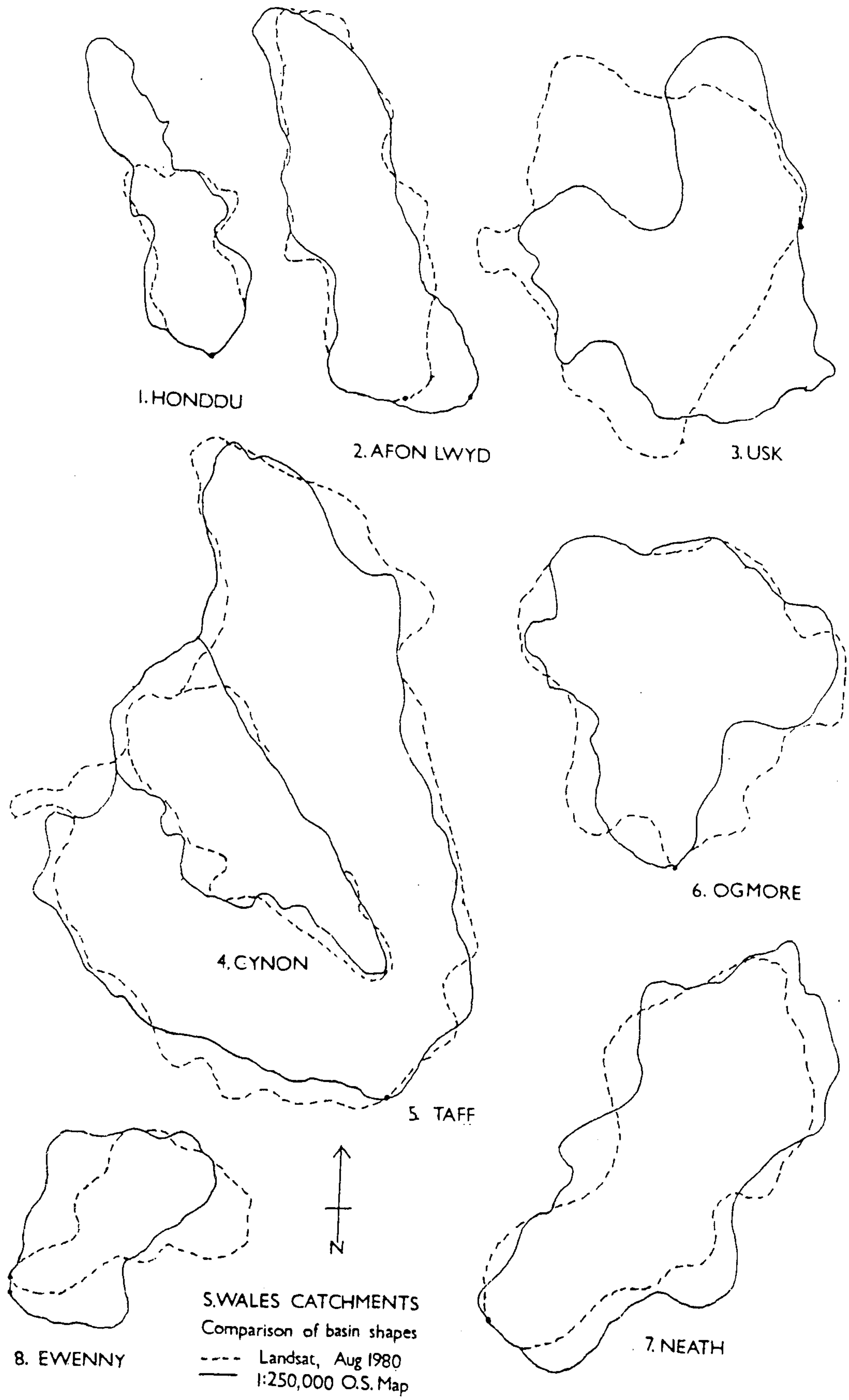


Fig. 3.8 Errors in catchment boundaries

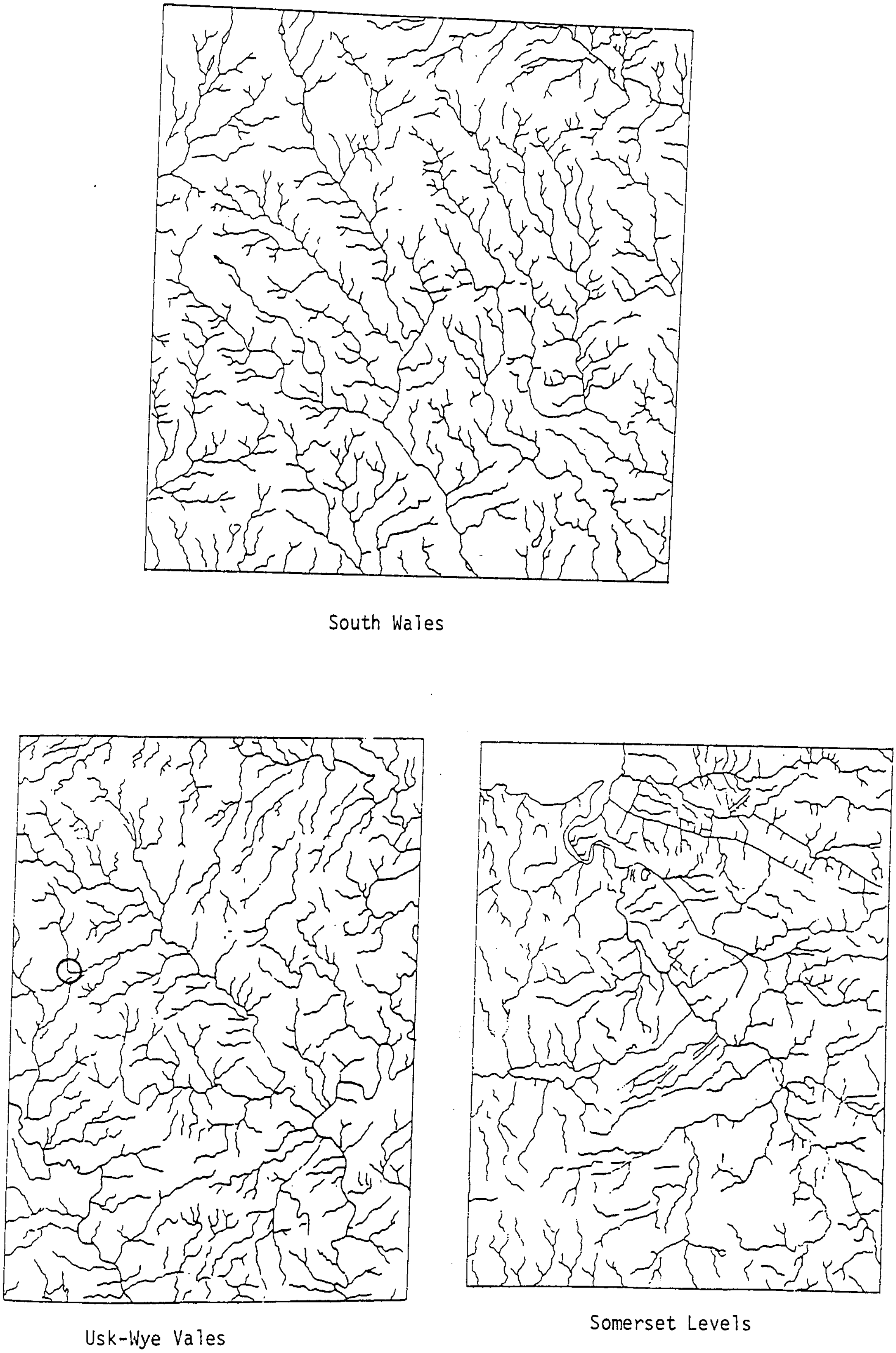
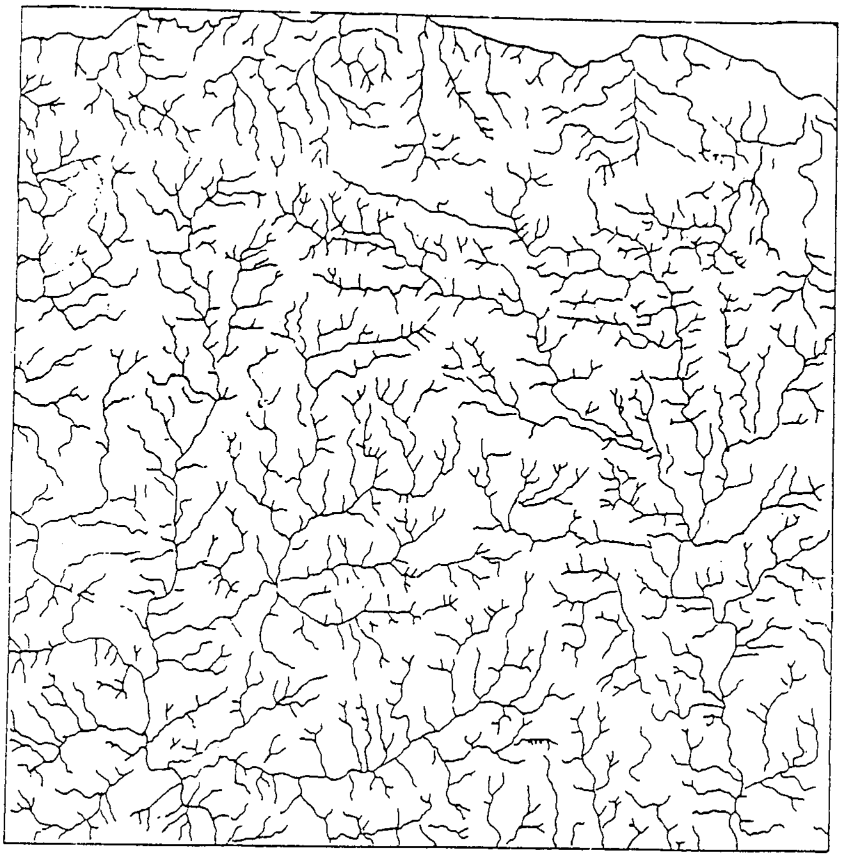
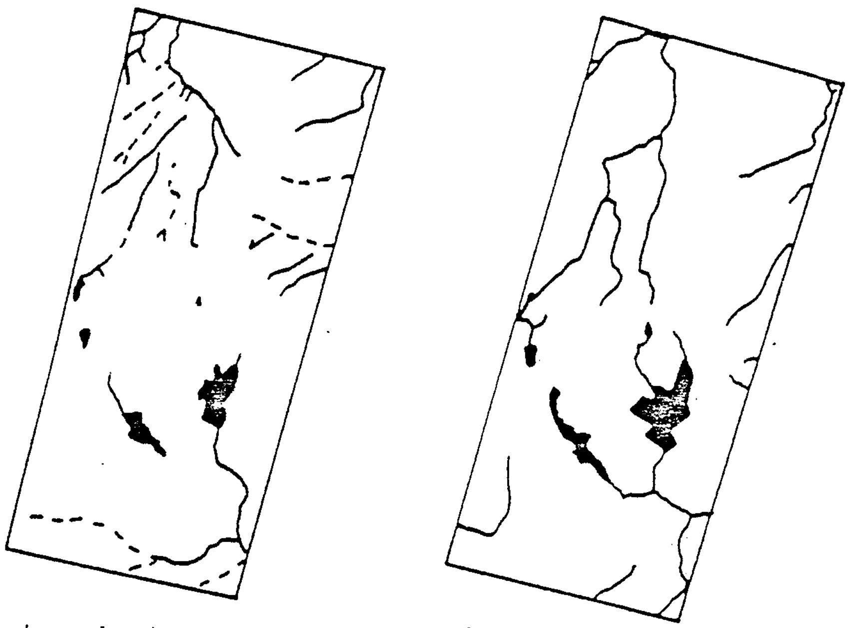


Fig. 3.9(a) Drainage network mapped from 1977 image



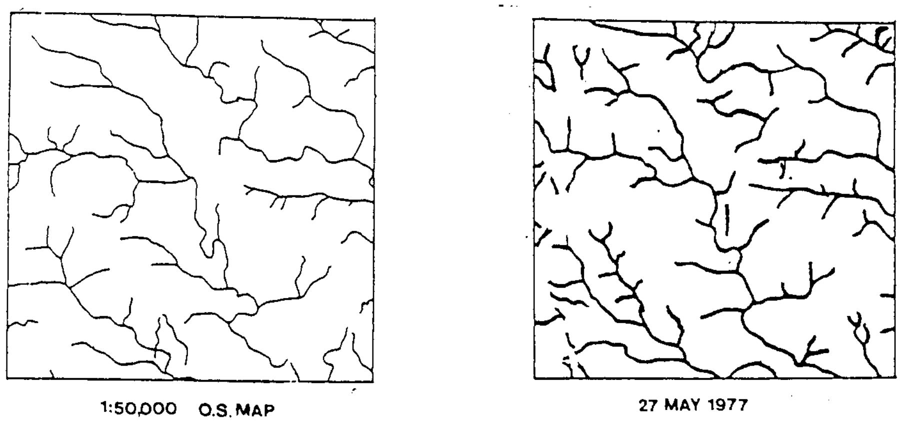
Exmoor



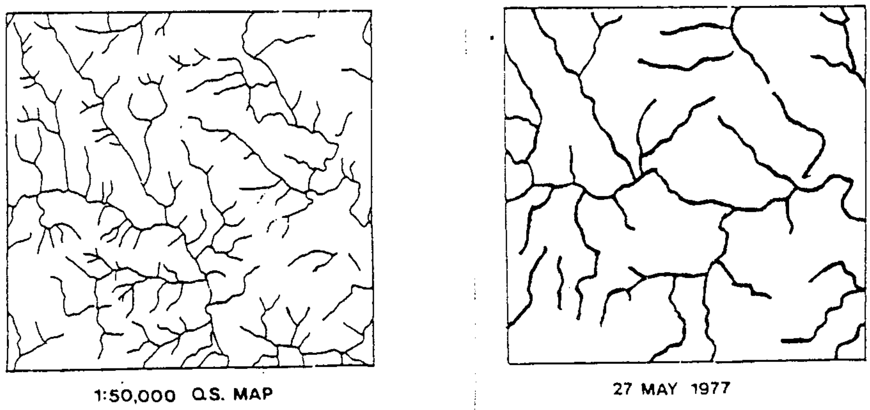
Landsat O.S. Map

Denbigh

FIG. 3.9(b) Drainage network mapped from 1977 image



Exmoor (Grid Square SS83)



Usk-Wye (Grid Square SO41)

Fig. 3.10 10km square samples for evaluation of correctness factor

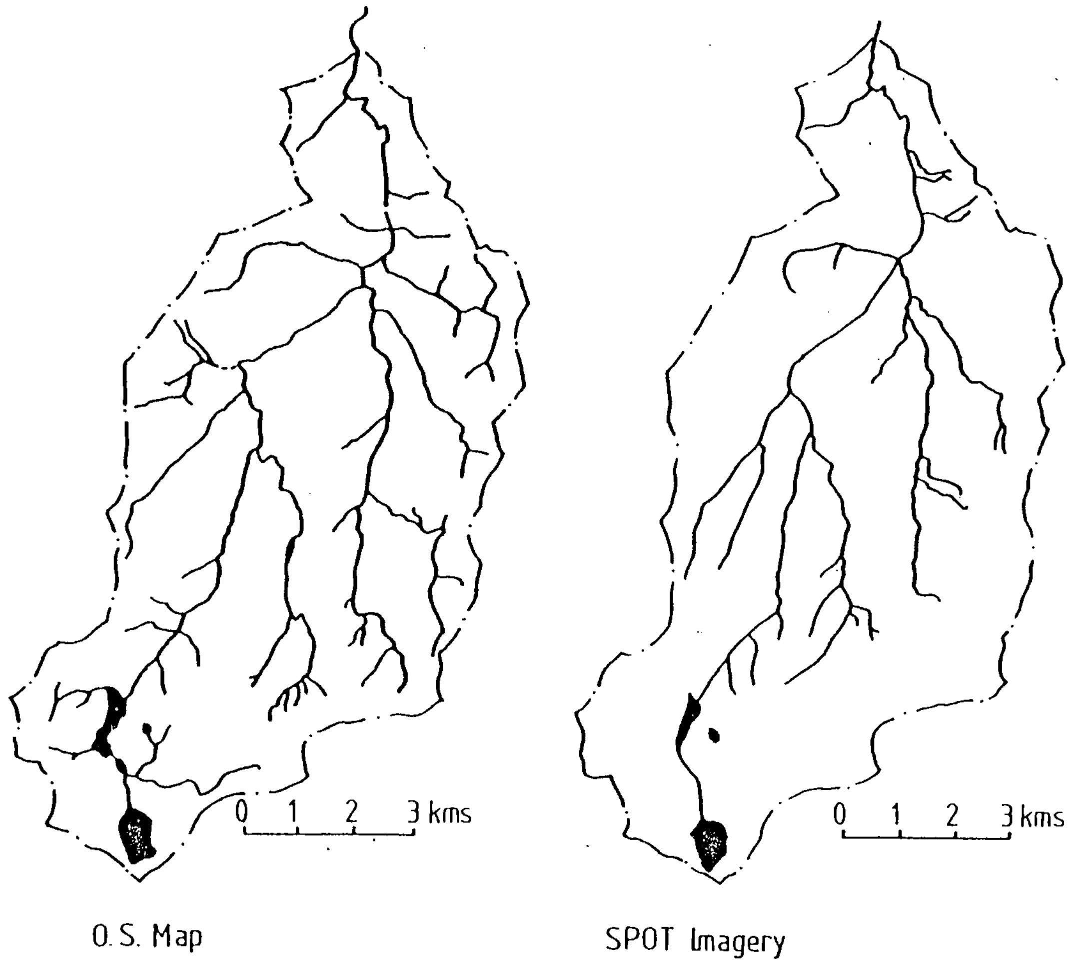


Fig. 3.11 Drainage mapping at 1:50,000 scale, Afon Aled, Denbigh.

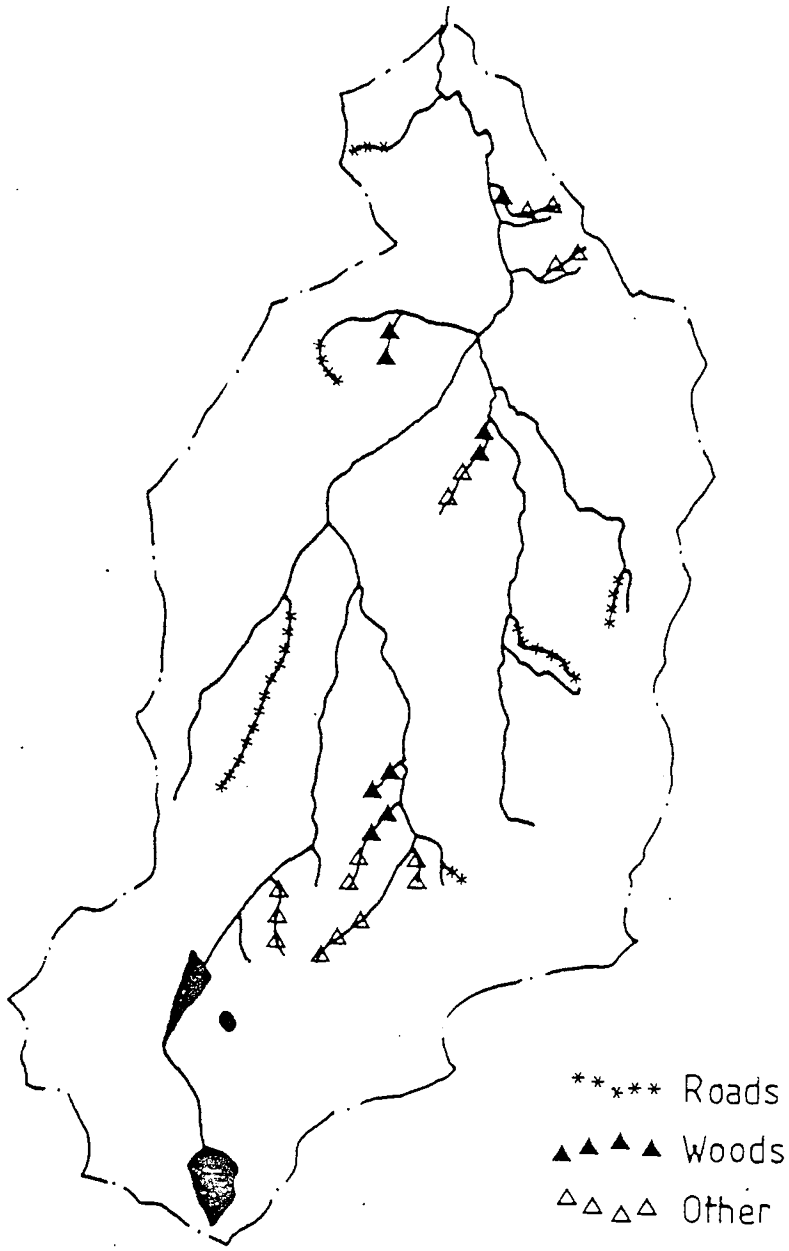


Fig. 3.12 Incorrect interpretations of Afon Aled drainage network.

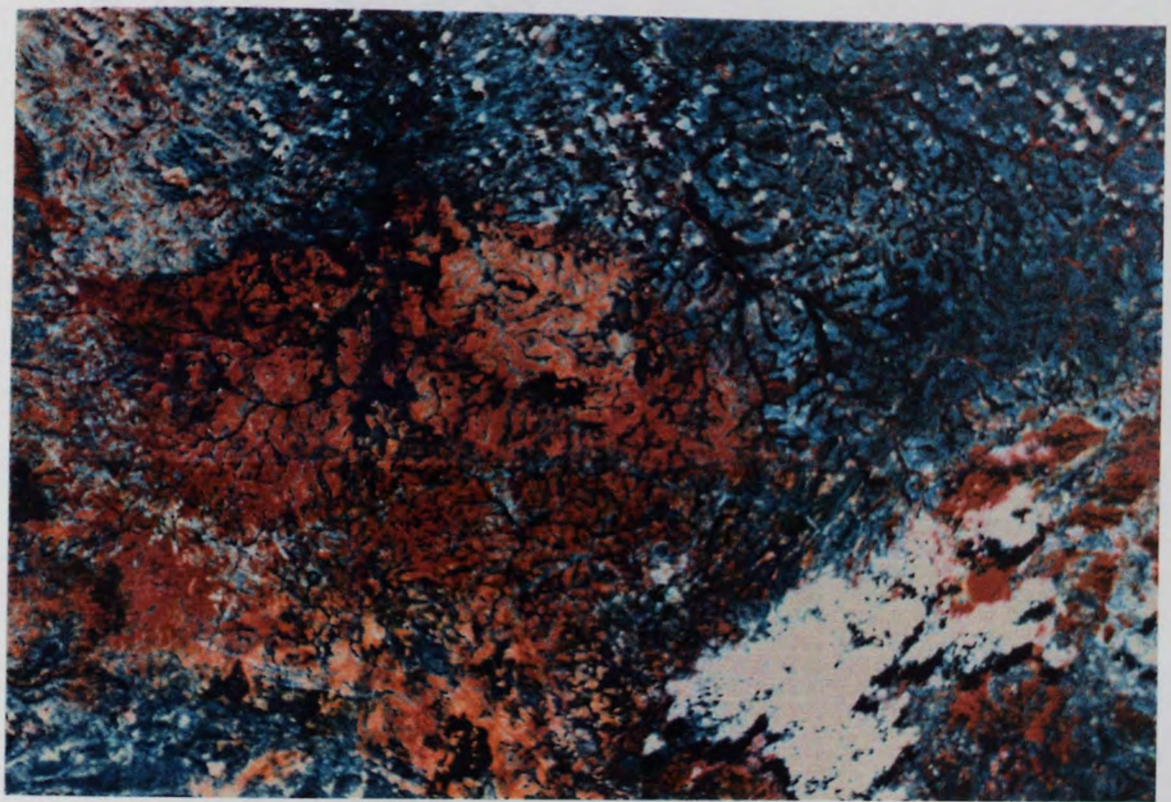
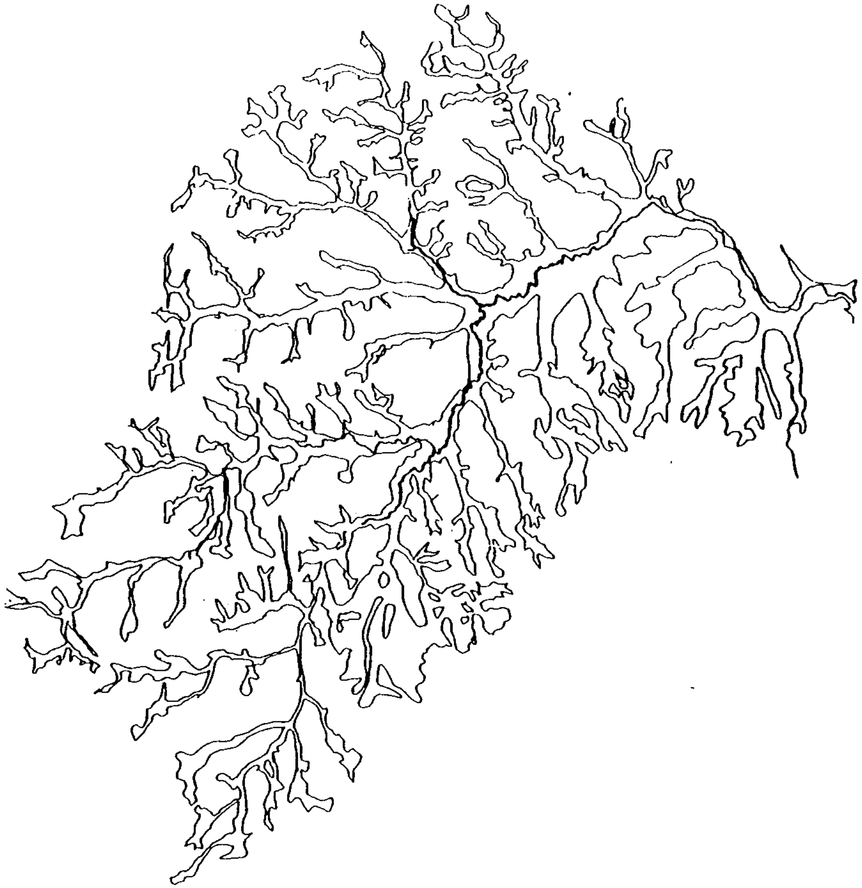
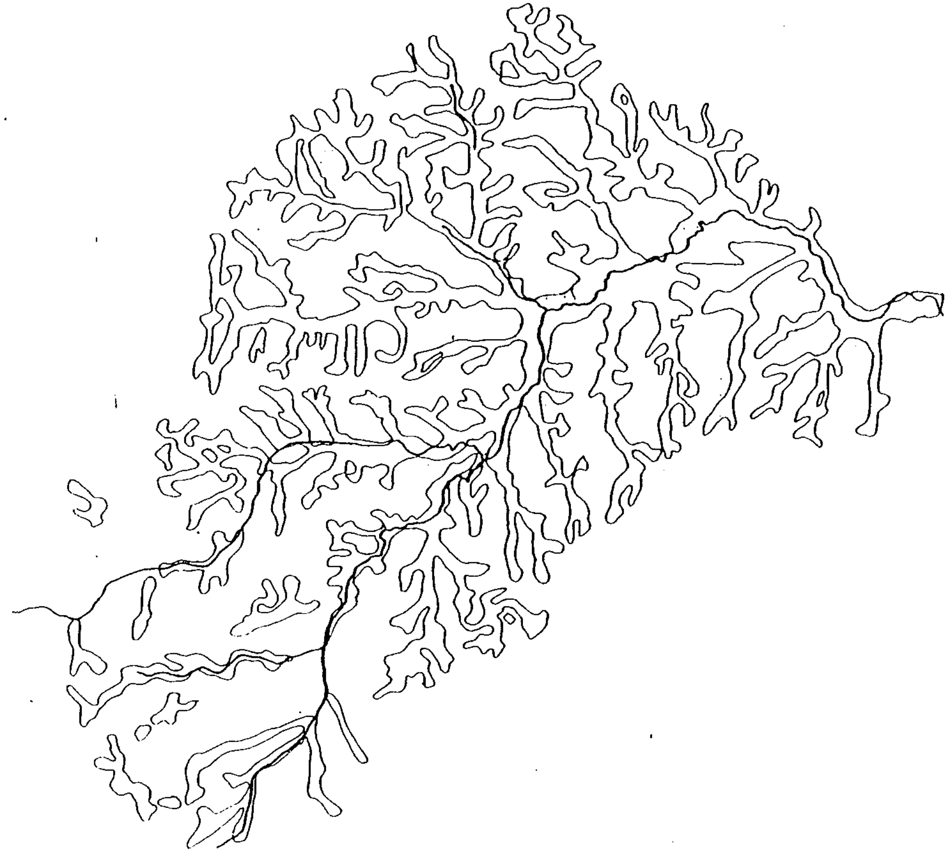


Fig. 3.13 Dwangwa and Rusa drainage networks

Landsat image 181/89 Malawi Nov. 1981

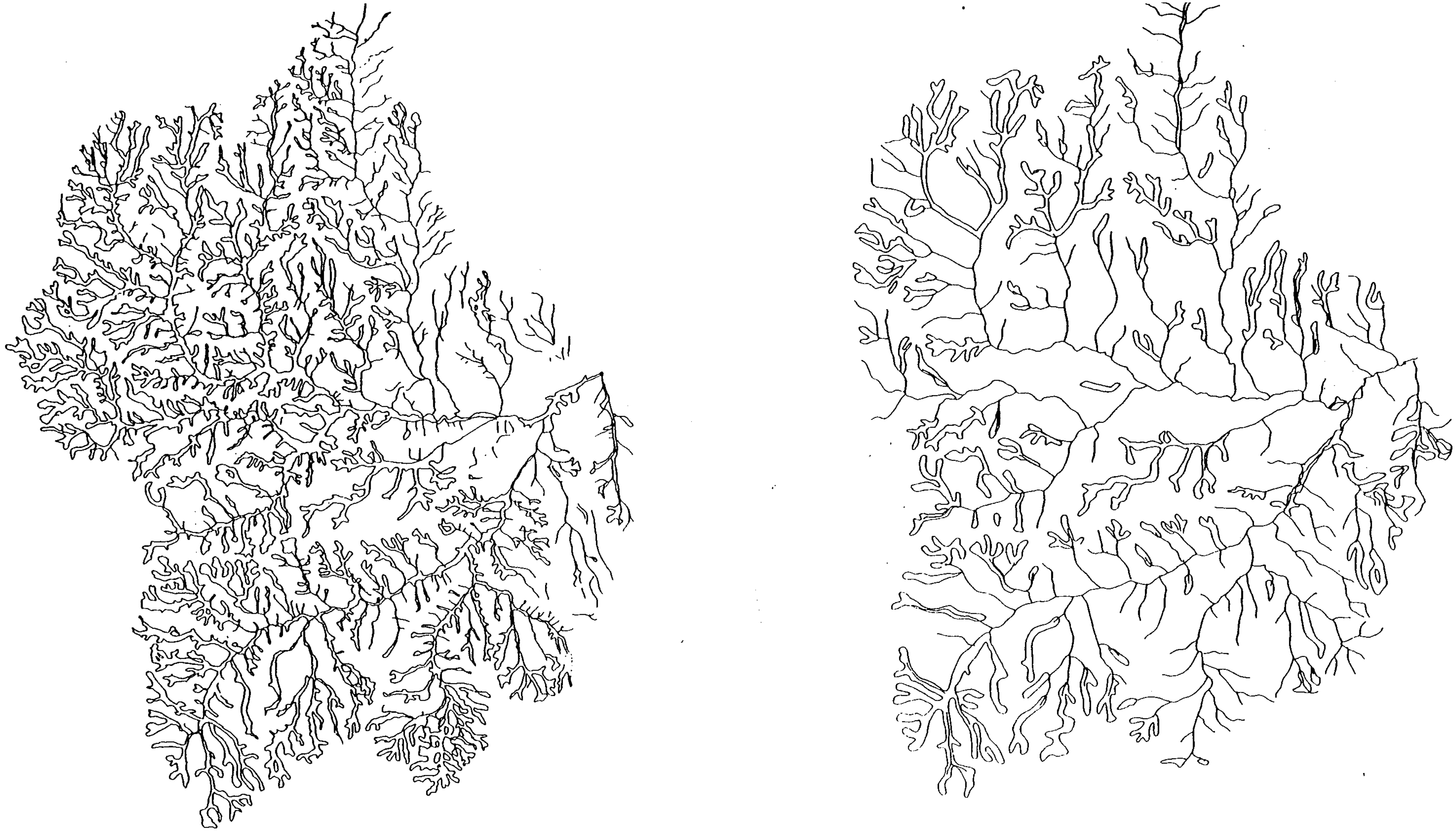


LANDSAT MSS 1:250,000



1:250,000 SCALE MAP

Fig. 3.14 COMPARISON OF LANDSAT INTERPRETATION AND CONVENTIONAL MAPS, RUSSA DRAINAGE BASIN



LANDSAT MSS 1:250,000

1:250,000 SCALE MAP

Fig. 3.15 COMPARISON OF LANDSAT INTERPRETATION AND CONVENTIONAL MAPS, DWANGWA DRAINAGE BASIN

Fig. 3.16 COMPARISON OF DRAINAGE NETWORK AND EXTENT OF DAMBO INTERPRETED FROM LANDSAT IMAGERY AND FROM CONVENTIONAL MAPS

10 km SQUARE 858 500, DWANGA BASIN

FROM 1:50,000 SCALE MAP

FROM 1:250,000 SCALE IMAGE

FROM 1:250,000 SCALE MAP

PART 3 MODELLING STREAMFLOW USING REMOTELY SENSED DATA

PART 3 SYNOPSIS

A report is given of the use of remotely sensed data in providing quantitative estimates of the surface water resource in a region, concentrating on simple models in which streamflow is estimated directly by a regression on catchment characteristics, or via rainfall-runoff relationships in which the parameters are estimated from catchment characteristics.

The rationale for using remotely sensed data rather than map data is presented in Chapter 1, which is followed by a review of the literature, based on a selection of important streamflow models.

The following chapters describe the development and testing of models for the prediction of streamflow in the South Wales area, which are robust enough to be re-calibrated for other regions.

Conclusions are drawn regarding the value of remotely sensed data.

PART 3 CHAPTER 1

INTRODUCTION

The argument which is examined in this section is that the spectral characteristics of the ground can be used as parameters in hydrologic regression models, to provide more efficient models than those based on conventional characteristics. The logic for this premise is as follows :

1. The parameters in regression models are chosen for their value in highlighting the differences that occur in observations i.e. they should be capable of explaining the variance in the independent variable. Causal relationships are not necessarily important in regression models e.g. the magnitude of the flood-producing rainfall need not enter the regression model for the 100 year flood.

2. The spectral characteristics of the surface may be encapsulated in a few features which vary widely from place to place (i.e. the spectral signature).

3. The spectral signature of a point is determined by the topography, soils and vegetation of the point, which are the dominant factors in determining the hydrologic response at that point.

4. Thus the many factors which determine hydrologic response may be replaced by the few factors needed to describe the spectral response, and a regression model may be developed using a few parameters which efficiently explain the variance in independent variables.

This approach would lead to a class of regression models which may be useful in an area for which models had not previously been developed, where conventional mapping was not available. An alternative would be to use the relation between spectral characteristics and physical characteristics simply to map the features that have been found to be useful in existing models i.e. conventional catchment characteristics. The benefit here is that parameters may then be estimated for areas where maps are out of date. The rapidity of change in the landscape due to agriculture, forestation and urbanisation is so great that mapping organisations are unable to keep up with the change except at very small scales, and remotely sensed data provides a valuable alternative.

Thus the investigations reported in this section are threefold :

(i) a correlation analysis to confirm that hydrological features are related to spectral properties,

(ii) the development of models based on spectral properties,

(iii) the development of models based on conventional catchment characteristics, which may be interpreted from spectral properties.

It was decided to limit the investigation at this stage to two simple types of model: regression models where streamflow is estimated directly from a regression on catchment characteristics, and rainfall-runoff models where streamflow is estimated as a direct proportion of rainfall, with the proportion being decided on the basis of catchment characteristics. The choice of spectral characteristics was determined largely by the preliminary correlation analysis. The choice of conventional catchment characteristics depended, though, on which existing models were considered to be conventional.

In Britain the most widely used models for estimating streamflow from ungauged catchments are derived from the technique developed by the Institute of Hydrology in the 1970's, which are best represented by the Flood Studies Report model for the mean annual flood (NERC, 1975) and the Low Flow Studies model for the 95 percentile flow (Institute of Hydrology, 1980). These are both straightforward regression models based on a set of parameters which includes geomorphology, climate, soil properties and land-cover.

In the USA the SCS Curve Number model is ubiquitous (SCS, 1972), forming the basis for simple flood estimation, and performing the rainfall-runoff function in more complex models. In this model the volume of flood runoff is estimated from a specified rainfall using a runoff coefficient, called the Curve Number, which is calculated on the basis of land-cover and soil-type, modified by an arbitrary choice of antecedant conditions.

Taking these three as being typical of simple stramflow models, the choice of "conventional" characteristics thus become : soil-type, climate, geomorphology and land-cover. Of these, land-cover is the most amenable to quantitative, objective interpretation from satellite imagery. Of the geomorphological factors, catchment area and drainage density have been discussed in Part 2. Stream slope cannot be estimated from Landsat imagery, but possibilities exist for the use of SPOT data, and for blending satellite data with digital terrain models. In the work which follows, therefore, the concentration is on the use of land-cover data in the two types of model. The estimation of soil conditions and climatic factors is discussed later in the chapter on recommendations for future work.

Summarising then, this section is concerned with the development of regression models for streamflow, based on catchment characteristics, where the characteristics may be spectral properties, or may be conventional characteristics which can be estimated from spectral properties. The sequence of operations to carry out this task were thus :

1. Select a set of gauged catchments within a hydrologically homogeneous region,
2. Examine the hydrological record for those catchments, and assess which features of streamflow are most amenable to regression modelling,
3. Acquire remotely sensed data, and assess which spectral characteristics will be most useful in regression modelling,
4. Examine the efficiency with which land-cover may be interpreted from remotely sensed data,
5. Examine the correlations between streamflow and spectral data, and assess how useful spectral data may be in explaining hydrologic features within the study region,
6. Develop regression models for predicting streamflow in ungauged catchments in the region,
7. Propose models which would be useful in other regions, where sufficient data exists for their calibration.

In carrying out this work, due regard was given to the availability of conventional data such as climate and topography from other sources, and data which could be extracted from satellite imagery using the techniques examined in Part 2 i.e. the geomorphological characteristics of the catchments.

The study area for this work was that part of South Wales lying in and around the coalfield i.e. the Usk, Taff and Neath basins, which has been described in Part 2.

Missing page(s) from the bound copy

PART 3 CHAPTER 2 REVIEW OF LITERATURE

As has been noted earlier, British regression models do not rely greatly on parameters which quantify the type of land-cover, whereas, in contrast, the most widely-used American models for flood estimation are strongly based on land-cover. Research into the applications of remotely-sensed data has been promoted largely by NASA, and has responded to the needs of large government agencies such as the US Department of Agriculture, and the US Corps of Engineers. Thus it is scarcely surprising that the literature on the use of satellite land-cover data in regression models for streamflow is almost entirely American in origin.

Following the launch of Landsat-1 in 1972, great interest was shown in the capability for distinguishing between various types of crop. The strategic value of this kind of information soon led to the development of powerful and reliable techniques for the classification of land-cover from Landsat imagery. Environmental scientists were quick to appreciate the value of timely, detailed maps of land-cover, and began to assess their value in models of environmental processes. In the realms of hydrology, the usefulness of satellite data to estimate the degree of impervious cover was recognised, and as early as 1975 Ragan and Jackson had reported on a comparison between Landsat interpretation and conventional air-photo interpretation for the assessment of impervious cover in the Anacostia River basin, which is a suburban catchment in Maryland. Low-level aerial photographs were used to categorise each point in a 0.45 ha grid cell arrangement into one of nine land-cover classes. The area in each class was then aggregated, and compared to the area estimated using Landsat imagery. Each category was similar to within 3%, and when an impermeability factor was calculated the Landsat value was 25.1%, compared with 23.5% for the conventional approach.

In reporting these very good results no mention was made of the probable difficulties in classification where high quality ground-truth does not exist, nor was there any discussion of the problems to be encountered when extending the classification to areas for which ground-truth is not available. Recognizing the potential errors that could be involved in land-cover estimation, Salomonson et al (1975) made a sensitivity analysis to determine what change would be brought about in a streamflow estimate due to a change in the land-cover estimate. They used a modified Stanford Watershed Model IV for three diverse catchments, and found that of the 26 parameters used in the model three related directly to land-cover (i.e. impervious area, forested area and water area), and that another two could be estimated indirectly from land-cover (volume of interception and hydraulic roughness). They found that it required as much as 20% change in any of the three areal factors to bring about a 5% change in runoff, and that runoff was even more insensitive to changes in the other two factors. Thus the conclusion was that Landsat imagery, even with the coarse resolution available at that time, could give land-cover data of an accuracy which was acceptable for estimating inputs to the model.

In a simple regression approach, Hollyday and Pluhowski (1976) examined the improvement that could be achieved in streamflow estimates by the addition of land-cover types to the list of catchment characteristics. The study was made for 39 catchments in Delaware, and it was found that four simple land-cover parameters (forest, field, urban and water) when added to information on soil-type, rainfall etc. could give a 20% improvement in the prediction of July streamflows. Other measures of streamflow, such as the 50 percentile flow were also improved, but to a lesser extent.

Voss et al (1978) made a regression study for flood flows on small catchments, and found that the inclusion of certain satellite-derived land-cover parameters gave decrease in the standard error, compared with the regression of flood flow on area alone. However, no consistently good data set was found, and the change in standard error was actually quite small, e.g. from 48% to 43%. Nevertheless, techniques for handling Landsat data were established, and the benefits of grid-cell data management were recognized.

Thus, the foundation was soon laid for either conceptual or regression modelling of all types of streamflow using land-cover estimates derived from Landsat interpretations. The main thrust of the American work, however, centred around the use of Landsat data for estimating flood flows in ungauged catchments.

The Fourmile Run studies

Following Ragan and Jackson's pilot study on the Anacostia basin, the work was extended to examine the usefulness of satellite data in urban flood models, such as the popular STORM and WREM models. The US Corps of Engineers' Surface Treatment Overflow and Runoff Model (STORM) is a continuous simulation model that can be used for flood estimation. At an early stage in the model rainfall quantities have to be translated into runoff volumes, which conventionally required the use of either the SCS curve number method or an empirical relationship based on data from recorded floods.

Jackson et al (1976) developed an alternative technique for calculating runoff quantities which was based on Landsat interpretations, and tested the technique on the Fourmile Run basin in Virginia (Jackson et al, 1977). The impervious area of the basin was estimated to be 39% which compares well with the 35% estimated by conventional means. In a simulation of the largest historic floods in the basin, the STORM model gave a standard error of 17.2% using Landsat data, which was only 1.6% higher than that using conventional techniques.

The Fourmile project was extended to examine the use of a much more detailed model, the Water Resources Engineer's Model (WREM), which split the basin into 179 sub-basins. Landsat data was used to estimate the percentage areas in each of six land-cover categories. The Landsat method was found to give a peak discharge for the whole catchment only 7.4% different from the conventional analysis (Jackson et al, 1977). The difference increased in the case of sub-basins, and the conclusion was made that the error using satellite data of that resolution (60m x 80m) would be unacceptable in catchments of less than several hundred hectares.

<u>The SCS curve number model</u>

A very widely used model for estimating flood flows is that developed by the Soil Conservation Service of the US (1972), which was originally developed for rural watersheds, but has been modified for use in urban areas (USDA, 1975). All SCS models are based on the manipulation of a volume of flood runoff, which is estimated from the design rainfall, using a runoff coefficient known as the curve number. The curve number is defined in terms of land cover and hydrologic soil class, and modified to take account of antecedent conditions.

An interesting early approach was made by Blanchard (1975), who attempted to estimate curve numbers using only the spectral characteristics of the catchment. Using 10 study catchments and 10 trial catchments all from one basin in Oklahoma, he derived regression equations for curve number, based on a variety of band combinations. Although good correlations were established, he concluded that his technique would be appropriate only in dry conditions with dormant vegetation, and would need recalibration for any cropping patterns which were different from the ones which he had studied. Therefore it had very limited applicability.

A more conventional approach was adopted by Ragan and Jackson (1980) for the Anacostia basin, in which Landsat imagery was used to estimate land-cover, which was then combined with information from a soils map to give the curve number. Insignificant differences were found between flood estimates using conventional survey, high level aerial photopgrahy and Landsat interpretation. An important issue raised by this study was the need to modify the SCS land-cover categories to comply with the classes that may be interpreted from Landsat imagery. This actually involves a very large reduction in the number of classes, down to the following five : Forest, Grassed open space, Highly impervious (e.g. industrial, parking), Residential and Bare ground.

Slack and Welch (1980) conducted a similar study, but for a large, rural catchment, in which they used only four hydrologically important classes : agricultural vegetation, bare ground, woodland, and open water.

The curve number for six sub-catchments were computed to within 2 curve number units of the values assigned by conventional techniques.

Jackson and Rawls (1981) in a study concerning curve numbers for urban areas made the important point that differences in land-cover between Landsat interpretations and conventional maps may not necessarily represent an error in the interpretation, but may actually be due to changes in land-use that have occurred since the production of the map. Clearly the same would be true in rural areas where forestation and changing cropping patterns could cause significant changes over the space of say 10 years.

The HEC-1 model

Following the work done by Ragan and Jackson, NASA commissioned a project to assess the usefulness of Landsat data in hydrologic work of an operational status. Working with NASA and the Corps of Engineers, Rango et al (1983) firstly examined the accuracy of an unsupervised classification of Landsat data, and found only 5% error at the basin level, but a 36% misclassification when examined at cell-by-cell level.

These data were used in the Corps of Engineers HEC-1 model, which calculates runoff using either the SCS curve number, or using Snyder's unit hydrograph with an estimated percentage imperviousness. The model is calibrated using regional data to provide a flood frequency curve. Rango et al found that there were insignificant differences in the frequency curves generated from conventional land-use data and the Landsat interpretation. The reason for such a close correspondence in results is probably due to the fact that the flood estimate is largely determined by the choice of antecedent condition (SCS) or the percentage runoff (Snyder), and that the influence of land-cover is relatively insignificant.

The most important result of this work is the recommendation made for the use of grid-based data management systems, in which hydrological data can be set alongside many other types of environmental data to provide a greatly enhanced richness of analysis. i.e. in a geo-information system.

Geomorphic parameters

In contrast to the effort that has been made to investigate the usefulness of Landsat land-cover estimates, very little work has been done on the use of geomorphic factors. Rango et al (1975), concluded that various geomorphic measurements from Landsat 1 were generally comparable to measurements derived from maps regardless of study area, and went on to report work then in progress which had initially shown that four physiographic parameters used in the Stanford Watershed Model could be suitably extracted from Landsat data for watersheds as small as $78km^2$. They further surmised that physiographic and land-use information from Landsat is compatible for use in general mean annual runoff predictions, but gave no details.

Killpack and McCoy (1981) working at the same scale of 1:100,000 extracted data on basin area, length of stream network, length of main channel and basin perimeter. They found a very strong cross-correlation between these variables and so chose just one, which was basin area, to regress against runoff. Using data from 12 basins within one hydrologic region, they found a good correlation between map data and Landsat data, and a good correlation between mean annual discharge and area. Putting these together they presented a regression equation for mean annual discharge using the Landsat derived catchment area, which had a correlation coefficient of 0.92. However, this cannot be taken as a particularly searching examination of the value of remotely sensed data, and is more a demonstration that average runoff per unit area does not vary much throughout their study region.

Summary

The bulk of the published work is concerned with the estimation of parameters for existing models, and stems from the USA, where techniques developed rapidly for the classification of Landsat images, and where the more common hydrologic models are based primarily on land-cover and soil-type. Very little interest has been shown in incorporating geomorphic parameters, such as those used in the UK Flood Studies Report.

The consensus is that Landsat interpretation can provide an accurate estimate of land-cover, provided that the range of classification remains small. Equally, the simpler hydrologic models work well with a limited range of land-cover types, and the inaccuracies incurred through Landsat interpretation are far outweighed by other estimates that have to be made, such as the antecedent conditions.

More complex models, such as the Stanford Watershed Model are relatively insensitive to errors in land-cover, and would work well with Landsat data. No conceptual model as yet has been designed to work solely with remotely-sensed inputs, but Ragan's team at the University of Maryland are currently developing appropriate algorithms. (Groves and Ragan,1983).

PART 3 CHAPTER 3 HYDROLOGY OF STUDY AREA

3.1 Selection of data

The choice of stations for study was governed by several criteria. Firstly, for successful regression modelling the data should cover a wide range of hydrological types, from flashy mountain catchments to lowland agricultural catchments, with an equally wide range of climatic inputs. Each station record was required to include good quality daily flow data over a long period of years, together with estimates of catchment rainfall. Where large abstractions or discharges occurred the record should be corrected to give 'naturalised' flows. Ideally, reliable flood estimates should also be available, but in practice it was found that a slightly different set of stations was needed for the flood modelling. There was also an important, but non-hydrological criteria, which was that all catchments should lie within one Landsat scene, so that there would be no problems with the calibration of spectral data.

A total of 15 stations were finally chosen, and the catchments are shown in Fig. 3.1. Data were acquired from the UK Surface Water Archive at the Institute of Hydrology. These were in the form of monthly rainfall and runoff, expressed as mm over the catchment, and the hydrographs for the 15 catchments are shown in Appendix 1. Flood data and low-flow statistics were collected from the records of the appropriate divisions of Welsh Water Authority.

An important feature of the part of the study was to model long-term average flow statistics, and a decision had to be made on the period to be taken for this average. If the average was made over a standard period, then it would have to be very small, as shown in Fig. 3.2, where it can be seen that the only period which is common to all records is the one year overlap (1973) between Usk (Trallong) and Sirhowy. Clearly a one-year period would not give a stable value for the station averages. An alternative would be to abandon the Usk (Trallong) record, which would then permit the use of the period 1973-79, but this is still a very short period of record for hydrological averages.

Assuming that the estimate of the average improves with the number of values used in its calculation, then the best estimate of an average flow is that for the Ebbw River, where 27 years of record gives an average runoff of 1039.0mm. The estimate of the Ebbw average for the period 1973-79 is 872.3mm, which shows a sampling error of 16%, and it can fairly be assumed that a similar sampling error would be incurred in every other record. (The error is probably so large because the period includes the 1975-76 drought).

This may be compared with the sampling error that would be incurred by taking the whole period of record for each station. This was examined by taking the largest record (the Ebbw River) and calculating the sampling error that would be incurred by using a period of record equivalent to those of the other stations (Table 3.1). The estimates are all low, reflecting the declining runoff of the 1970's, but they are all considerably more accurate than the 1973-79 estimate.

Thus, on the one hand we have the consistency offered by the use of a standard period, but a poor estimate of the parameter which is actually of interest. On the other hand we have better estimates of long-average runoff for each station, but with considerable variation between stations.

The decision was made to use the whole period of record for each station, on the grounds that it provided a more robust estimate of long-average runoff, and that it used all the available data. In making this choice it was accepted that the data set used in the regression analysis would not be homogeneous.

With regard to annual maxima, no attempt was made to analyse flood series, it being assumed that an estimate of mean annual flood was sufficient, and that existing models for estimating the T year flood would be used. It was noted that the period of record for flood flows was very variable (Table 3.2) which leads to statistical bias, and that the period of record was generally short (10 of the stations with less than 20 years of record) which leads to sampling error. Also, according to the information listed in the Flood Studies Report (Vol. IV, Chapt. 3 and Para 1.2.5), only seven of the stations had gradings better than 'adequate' for the quality of flood-flow gauging. Additionally, three of the Welsh Water Authority records included estimated values for the largest flood in the series of annual maxima. Thus it was concluded that the estimates of mean annual flood were prone to error. Unfortunately, it is difficult to quantify the error involved, especially with the small number of station-years of record available. In the Flood Studies

Report (Vol. I 4.3.6) an attempt is made to examine measurement and sampling errors in a very much larger data set. Results are inconclusive, but it is suggested there that the use of short duration records makes little difference to the values of the coefficients calculated in regression models of the type used here. It is also suggested (Vol. I 2.5 and 3.1) that the sampling error for Mean Annual Flood would be of the order of 10-15% for records of about 10 years length. Thus, the decision was made to go ahead with all of the flood data, but to recognize that unquantifiable errors may exist in the results.

The Minimum Daily Flow and 95 percentile flow statistics were also taken directly from Welsh Water Authority records, with no attempt at analysis. Comments regarding sampling error and accuracy of gauging apply equally here, to which must be added the problems of artificial abstractions and discharges which strongly affect these low flows. Thus in the regression analyses which follow, the values of the dependent variables contain some error of estimation. It was assumed that there would be no consistent bias in estimation across all of the catchments, and that error terms would be Normally distributed, and that a Type I regression model could thus be applied. Nevertheless, it should be recognized that some part of the variance of the dependent variables cannot be explained by the correlation with the independent variables, and that the performance of the regression models will be limited by the errors of estimation.

3.2 ANALYSIS OF DATA

1. Summary of hydrological statistics

The monthly records were processed to give the long-average rainfall and runoff statistics shown in Table 3.2, and the mean elevation of each catchment was calculated from Flood Studies Report data (NERC, 1975). Mean annual floods were calculated as the arithmetic mean of the series of annual maxima from Welsh Water Authority records, while the low-flow statistics of Minimum Daily Flow and 95 percentile flow were taken directly from Welsh Water Authority records.

A full summary of hydrological statistics is given in Table 3.2 along with the definition of each variable used. The coefficient of variation (standard deviation/mean) is given for each hydrologic variable, from which it can be seen that there is a wide range in all of the parameters, as is desirable for regression modelling

2. Annual rainfall-runoff statistics

Annual totals of rainfall and runoff were computed for each catchment, and are tabulated in Table 3.3. It can be seen that runoff varies widely from station to station, and appears to be strongly linked to rainfall.

It was assumed that the annual water balance may be represented by the following equation :

Run off = Rainfall - Evapotranspiration
± Change in groundwater storage

Over a long period the groundwater out-flow will be matched by recharge, and the net change in storage will be zero. Thus the average runoff (LARO) will be simply the average rainfall (LARF) modified for evapotranspiration losses i.e.

LARO = LARF - Losses

This valuable, albeit, simple hydrologic model could be put to practical use if a method could be found for predicting the average evapotranspiration losses for a catchment. Since evapotranspiration is linked to land-cover and topography it may be possible to derive a regression model based on catchment characteristics, which would predict the losses term. Then, knowing average rainfall, it would be simple to calculate an estimate for long average runoff.

However, it was felt that a model for annual runoff totals would be more useful, since it could be used to provide a time series of runoff values as well as the long-average value. Firstly it was assumed that any change in groundwater storage over one year would be small. This assumption was made on the basis that there are no large, uniform, continuous aquifers in this region, and was supported by an examination of the lag-one serial correlation, which was found to be statistically not significant in any of the catchments. This implies that the response of the catchment to rainfall is mostly dissipated within one year, and hence that groundwater levels return closely to normal within that year. Secondly, the annual losses were regressed on annual rainfall, and no significant correlation was found in any catchment i.e the losses term was found to be independent of rainfall. Thus the water balance equation could be modified into a simple linear relationship :

ARO =-A + B.ARF

Where ARO and ARF are the annual runoff and rainfall, and the constants A and B may be found by linear regression. The results of a regression analysis are shown in Table 3.4, in which it may be seen that the correlation between annual rainfall and runoff is very strong and that the standard error of estimate for annual runoff is low, averaging 67mm. Thus, if this model were used to predict runoff, a 95% confidence limit of approximately 130mm could be expected, which is acceptable with annual runoffs in the region of 1000mm.

It was noticed that the values of the coefficient B were all close to unity. A statistical test was made which showed that at 95% confidence level the coefficient B was not significantly different from unity in any catchment. Thus the model may be simplified to :

$$ARO = -a + ARF$$

where a may be thought to represent the average annual losses due to evapotranspiration. Values of 'a' are given in Table 3.4.

This model has the benefit of being a more acceptable representation of the physical processes involved, and is less dependent on purely statistical relationships. The simplification leads to very little change in the error of estimation, except in the case of the River Mellte, where a large increase was found (Table 3.4). However, the average standard error increased to only 69.0mm, and so it was decided to use this simpler model.

In order to use this model in ungauged catchments, it would be necessary to make an estimate of the value of the losses term. Since losses are linked to the type of vegetation cover and the topography of the catchment it appeared that it would be useful to establish a regression model in which the losses term may be estimated from a set of catchment characteristics. Clearly these should include physical characteristics, such as height and land-cover, but for the reasons outlined above, it may be helpful to use spectral characteristics, which may additionally reflect the climatic characteristics of the catchment.

Turning to monthly models, there is a possibility of constructing rainfall-runoff models for each month, but it was felt that any model which required 12 parameters to be estimated from such a short period of record would be unreliable. If monthly estimates are required it is recommended that annual data be estimated, and then distributed according to the magnitude of the monthly averages.

3. PROPERTIES OF THE TIME SERIES

The annual rainfall-runoff model may be thought to work best where there is a reasonably rapid response-time, and where the groundwater influence on streamflow is small.

Looking at the hydrographs in Appendix 1 and the typical hydrograph drawn to larger scale in Fig. 3.3, it can be seen that the base flow component is quite high for several of the study catchments. This gave rise to the idea that errors in the rainfall-runoff model may have been caused by a carry-over of base-flow from one year to the next, in which case a simple Markov model of the following form would be useful :

$$ARO_t = r(1)\ ARO_{t-1} + \text{random component}$$

where r(1) is the lag-one serial correlation coefficient.

The analysis was restricted to those stations whose record was thought long enough to be used in a time-series analysis, and the lag-one serial correlation coefficients of the pre-whitened series are given in Table 3.5, where it can be seen that they are statistically not significant. Thus the Markov chain was considered not to be an appropriate model. Longer period correlation coefficients were also statistically not significant, as can be seen from the typical serial correlogram shown in Fig. 3.4.

Having found no serial correlation in the time series it seemed possible that a random noise model of the form

$$ARO_t = c_1 + a(t) + c_2\, a(t-1)$$

might be appropriate, where a(t) is a white noise function. Models of this type were fitted to the data, but their performance was very poor compared with the simple rainfall-runoff model.

Returning to this model i.e.

$$ARO_t = a + b\, ARF_t$$

the residuals from the linear regression were examined, and it appeared that there was some link between one year's residual and the previous year's runoff i.e. a large runoff one year would often lead to a large residual for the next year. Thus an attempt was made to improve the basic rainfall-runoff model by adding a serial correlation term viz.

$$ARO_t + a + b\, ARF_t + c\, ARO_{t-1}$$

This was tried for the eight catchments with long records, but there was no consistent improvement in the standard error of estimate for annual runoff.

4. CONCLUSIONS

Values have been found for hydrological variables for which models would be useful. The values for the 15 study catchments have a large variance, as is needed for regression modelling. A simple model for estimating annual runoff from annual rainfall has been found, whose parameters could be estimated from a regression model. The time series properties of the longer series have been investigated, but neither serial correlation nor white noise models work well, and no improvement can be found over the simple rainfall-runoff model.

PART 3 CHAPTER 4 CATCHMENT CHARACTERISTICS

4.1 SELECTION OF CATCHMENT CHARACTERISTICS

The choice of catchment characteristics to be used was governed by the two main aims of the study : to determine whether satellite image interpretation could provide useful estimates of parameters used in conventional models, and secondly to investigate whether spectral features of the scene perform better than conventional map-based parameters when used in models.

In conventional models catchment characteristics fall into four basic categories : land cover, soils, climate and geomorphology. The most important empirical model using land cover is the SCS model for flood estimation (SCS, 1972) and it has been determined by Ragan and Jackson (1980) in a sensitivity analysis that a Level 1 classification is adequate for urban studies i.e. a classification into 5 basic types of land cover : forest, grass, highly impervious, residential and bare ground. In Britain, the Flood Studies Report method for ungauged catchments gives regression equations which include 'urban' but no other land cover types, although it is stated that 'forest' was investigated as a possible parameter.

With regard to soils, it has been recognized that classification of soil types from satellite imagery is possible in conditions of sparse vegetation, but is difficult and unreliable in temperate regions where all soils are densely colonised by vegetation (Campbell, 1987). Thus, in this study no attempt was made to classify the soil parameters used in conventional models, even though it is recognized that estimation of the curve number in the SCS method depends greatly on soil type, and that the 'soil' parameter appears in all of the regional equations in the Flood Studies Report method.

Climatic factors such as rainfall statistics clearly cannot be estimated from imagery, although there is a possibility of inferring the 'state' of the catchment at the time of the image i.e. soil moisture, surface wetness etc. These properties may be useful in real-time forecasting, but are not relevant to the estimation of long-term statistics.

Some goemorphological features such as drainage density may be estimated with confidence from imagery, as discussed above, and are important in the British flood models. Others, such as stream slope, are used in models and may possibly be estimated using SPOT stereo imagery, but cannot be estimated from the Landsat imagery investigated in this study. (However, it must be recognized that they could easily be estimated from digital elevation models, and should therefore not be ignored).

With regard to the spectral properties of the catchment, there are many bands and band-ratios that could be used, but it is difficult to perceive,a priori, which spectral properties would have the strongest statistical link with streamflow. Band 7 (near infrared) was selected for study, because reflectance in this wavelength is known to be strongly linked with the presence of water on the surface or in the upper soil horizons. Examples of the spectral response of various soils in changing water conditions are shown in Fig. 4.1.

Two band ratios have been associated in the literature with vegetation conditions, which are likely to be related in some way to long-term water availability. These are the 'Vegetation Index' which is Band 5/Band 7 and the 'Biomass index' which is Bands (7-5)/(7+5). Consideration of the spectral response curve will also show that Bands 4/6 is related to the wetness of soils. Benefits of using ratios are that they are dimensionless, they are free from calibration problems, and they remove the effects of sun-shading (Sabins, 1978).

The principal component transformation described above has the effect of compressing the data into one new band, the first principal component, which then contains up to 98% of all the variance in the data. Therefore, even though it is impossible to ascribe any physical meaning to the first principal component, it is potentially a powerful statistical parameter.

The choice of catchment characteristics for the regression was thus made from : the basic land-cover types, elevation, slope and stream-frequency, and from Band 7, the Vegetation Index, Biomass index and the First Principal Component.

4.2 ACQUISITION OF LAND-COVER DATA

Ideally, a sufficient number of images would have been used so that seasonal and longer term changes in spectral properties could have been investigated. At the time of the study a total of 16 Landsat scenes of the area were available from the National Archive and no SPOT scenes were available. Unfortunately, when photoprints of the images were examined it was found that in the majority of scenes thick cloud or snow obscured the area of interest, and only 6 images were of practical value.

A detailed examination of imagery then showed only one image to be completely clear of cloud, and free from haze. This was the image of 27 May 1977, which was available geometrically corrected and transformed to a 50m pixel size. One other image was free of cloud but showed an apparently uniform slight haze over the whole scene. This was the image of 26 April 1984 and it was selected so that at least one alternative set of data was available for comparison (Fig. 4.2).

The images had been system corrected i.e. the systematic errors introduced by the imaging sensor, the spacecraft and the data transmission system had been removed. However, there remained non-systematic distortions, so that images were not geometrically correct and residual errors remained in the sensor-calibration, which appeared as the common striping effect known as sixth line banding. Clearly, when remotely sensed data is to be used in a geo-information system or for any quantitative study it is essential that images are geometrically corrected so that locations, lengths and areas can be measured correctly. In this case it was decided to transform the image data to the map projection used by Ordnance Survey i.e. the Transverse Mercator Projection. However it must be recognized that transformations cannot be made perfect, and that some residual error will remain, no matter how many control points are used. In the geometric transformations made on the images for this study the worst Root Mean Square error was 0.7 pixels i.e. at the 95% confidence limit, individual data values may be assigned to points which are within ± 1.96 RMS errors, or ± 1.4 pixels of their correct map location. For Landsat MSS data, which has a nominal pixel size of 60 x 80m, this means that 5% of the data may have locational errors in excess of 112m. The importance of such errors depends on the use to which the data is being put and also on the spatial variability of the data itself. For instance if the image is being interpreted to provide a drainage map, then an error of 100m in the location of a stream would be considered to be very important, whereas a similar error in the location of the boundary of land-use class may have very little effect on the overall proportions of land-use within a large catchment, particularly where land-use remains uniform over large areas.

With regard to sixth-line striping, there are algorithms which successfully remove the stripes. Unfortunately, to do so, all data values within the image are adjusted, and the effect is to smooth or blur the image, to an extent which depends on the severity of the striping. In the images used here the striping was not severe enough to prevent visual interpretation of drainage networks, and was not strong enough to interfere with the classification process. Thus, on the principle that it is better to avoid any unnecessary modification of the data no de-striping was carried out.

The problems of geometric and radiometric accuracy are rapidly diminishing as images from the second generation satellites (Landsat TM, SPOT) are becoming available. Greatly improved ground resolution and changes in sensor technology have combined to provide images of extremely high quality, which are now beginning to enter the national archive.

The alterations to the data caused by geometric correction were accepted as being necessary, but the decision was made not to attempt any radiometric correction, which might degrade the data even further. Thus a degree of sixth-line striping was accepted.

The initial choice of land-cover types for classification was based on the American literature, and was : open water, forest, grassland, arable land, urban and industrial. When carrying out the classification, though, it was noticed that there were two very distinct types of spectral signature in areas which had previously been thought to be uniform grassland. The two signatures are defined by the scattergrams shown in Fig. 4.3. The difference was so marked that a decision was made to continue the classification on the basis of two different grassland types, and test their significance in the regression models.

Urban areas were deemed to include suburban land, while industrial areas included large factory sites, quarries and opencast sites. However, a preliminary analysis showed that areas of urban and industrial land were small in all catchments, and that it was difficult to make a consistent distinction between the two, and that regression models worked equally well with a combined classification. Thus the decision was made to lump the urban and industrial classifications together and the final set used for regression studies were the following six categories :

Open water

Forest

Arable land

Grassland (1)

Grassland (2)

Urban

The classifications were carried out using the clustering algorithm of a commercial software package on a subscene of the image that was just large enough to include all of the study catchments. The sequence of operations is given in Appendix 3

The classified image of 27 May 1977 is shown in Fig. 4.4 and an example of the classification for one catchment is shown in Fig. 4.5. Summary statistics are given in Table 4.1, and the variation in land cover from catchment to catchment is illustrated in Fig. 4.6. Examining the statistics of this classification, as shown in Table 4.2, it can be seen that there is a strong variation between catchments, and very little cross correlation between variables, as is desirable for regression analysis.

The accuracy of the classification appeared to be good, but of course it was impossible to confirm it in detail against the land-use which pertained 11 years ago. This is an important drawback in the technique of land-classification, and casts a serious doubt on the results. One way of checking the classification is to repeat it using an image of different date, but similar time of year.

Fortunately, the other reasonably clear image which was available i.e. April 1984 fell into that category, and it was used to make a fresh classification. Firstly, the 1984 image was classified using exactly the same training areas as those used in the classification of the 1977 image. Surprisingly, the result was quite different from the 1977 classification (Figs. 4.4 and 4.7). The most marked difference lay in the distribution of the two grassland types, and so it was decided that a field study was necessary to investigate what ground conditions these spectral signatures actually related to. This is reported in the selection which follows.

The 1984 image was then classified using a fresh set of training areas, and a more similar distribution of grassland was achieved, but at the expense of a less precise demarkation of forest areas. Several more classifications were attempted, each resulting in a slightly different distribution of land-cover. The conclusion was that imagery of the best quality is needed for reliable land classification, and that the best quality ground-truth should be used, which is extremely difficult when dealing with images that are not contemporary. Given that these ideal situations would rarely be encountered in practice, it was decided to use the 1977, 1984(1) and the 1984(2) classifications in the regression study, in order to investigate how sensitive the models would be to differences in land-use classification.

4.3 INVESTIGATION OF GRASSLAND TYPES

The first impression of the upland areas from maps and from casual observation, was that they were uniformly grassland, predominantly used for rough grazing by sheep. However, the spectral signatures indicated some marked differences within the grassland areas, and this was supported by the importance of this distinction in the regression models (which follow). A field survey was thus organised, which was carried out in the Brecon Beacons National Park, in which plant species were identified and soil conditions noted on four transects that were planned to lie across the two grassland types identified by the spectral signatures but on different topographies.

Figure 4.8 shows the imagery of the area, marked with the transect lines and with an indication of the major geologic units. Transect 1 was across an area classified as Grassland (1). Vegetation along this transect is shown in Fig. 4.9, where it can be seen that the area is one of low gradients, in which limestone pavements are apparent, with mostly thin peaty soils, supporting a vegetation association of fescue, mat-grass and bilberry (Festuca ovina, Nardus stricta and Vaccinium myrtillus) commonly known as Nardus grassland. This association persisted through the nothern end of the transect where slopes were steeper, but the underlying rock was the Grey Grits of the Old Red Sandstone series. Locally, in hollows, conditions were wet, and Sphagnum bog replaced the grassland.

Transect 2 i.e. Grassland (2), crossed the headwater zone of the Afon Hepste, in which gradients were found to be very slack, and conditions were wet, supporting an extensive association of purple moor-grass (Molinia) and Sphagnum moss, with some occurrences of soft rush (Juncus squarrosus).

The third transect was made on the steep flanks of Cefn Crew, which was found to be well-drained soil overlying Brownstones, with predominantly sheep-grazed fescues on the lower slopes, grading to bilberry moorland on the upper slopes. Transect 4 was along the top of the ridge between the two main valleys, again on Brownstones, but this time with very gentle slopes. Drainage was found to be poor, and the conditions were wet, with almost uniform Molinia bog.

The conclusion from this investigation was that the Landsat classification had distinguished between the freely drained Nardus grassland, and the Molinia moorland in which drainage was impeded, and conditions frequently wet. Clearly this is a very significant distinction in terms of hydrologic processes, and so the two grassland classifications were retained separately in the regression analysis.

4.4 ACQUISITION OF OTHER CATCHMENT CHARACTERISTICS

Band 7 values were taken directly from the May 1977 data set. The various band ratios were calculated, using specially-written programmes which applied the necessary offsets and re-scaling to give values which lay within the range of the equipment being used i.e. between 0 and 255. The principal component transformation was carried out using commercially available software. A masking operation then provided the set of data for each of the study catchments.

Having constructed a detailed map of spectral values in each catchment some summary statistics had to be chosen for the regression. The decision was made to use the mean value from each catchment, even though it was recognized that a great deal of information concerning the spatial and statistical distribution of data would be wasted, as is the case with any lumped model.

Tables of value for each spectral band, and for the selected band ratios are given in Appendix 2. It may be seen that the values of bands 4 and 6 do not have a very great variance, which confirms the decision not to use these in the regression analysis. Band 6 and Band 7 are highly correlated ($r^2 = 0.997$), which confirms that it is appropriate to use only one single band in the analysis.

It would not be correctto base any decisions on the mean or variance of the spectral ratios, since these are entirely artificial, and were determined by the arithmetic routines used to calculate them. However, scaling and spreading do not affect the covariances which exist, and it is informative to examine the cross-correlation matrix, (Table 4.3) from which it may be seen that the band ratios are highly cross-correlated.

There are several reasons for this. Firstly, in the Principal component transformation the data are not changed in any fundamental way. Each data point in n-dimensional space remains fixed in position relative to all other points; it is only the position of the axes that changes. Taking the simple 2-dimensional case as an example, it can be seen from Fig. 4.10(a) that the Band 1 data is strongly correlated with Band 2. Rotation of the axes (Fig. 4.10(b)) does not alter the data in any way, but the values of PC1 are now de-correlated from PC2 (Fig. 4.10(c)). This is the purpose of the transformation i.e. to compress the information contained in the data into one new band, PC1. However, it can clearly be seen that the values along PC1 are in a strict geometric proportion with the values along B1. Therefore PC1 must be 100% correlated with B1.

Thus, in techniques involving visual interpretation there is benefit to be gained by transforming the data, because careful choice of transformation can yield a single band which contains over 95% of the information in the scene. However, the conclusion reached here is that there is absolutely no benefit in transforming the data for a regression model, because the transformed data will be 100% correlated with the original data.

In the case of the band-ratios, it is clear that the vegetation index, which is Band 5/Band 7, must be very strongly correlated with both Band 5 and Band 7, and, since there is strong correlation between all the single bands, these correlations will extend to all combinations.

Turning now to the geomorphic variables, i.e. elevation, stream slope and stream frequency, the values used were the generally accepted values taken from the Flood Studies Report. This was on the assumption that such data would normally be extracted from maps or from digital terrain models to give "true" values. Tables of values of geomorphic parameters are given in Appendix 2.

PART 3 CHAPTER 5 REGRESSION MODELLING

The aim of the regression analysis was to develop models for the Mean Annual Flood, the 95 percentile flow, the Long Average Runoff, the Minimum Daily Flow and the losses term of the annual rainfall-runoff equation. Clearly, the flow statistics would be strongly related to catchment area, and so for the regression analysis, these were replaced by the flow per unit area. The choice of independent variables was thus :

The coefficient of the annual rainfall-runoff model (a)

Mean annual flood per unit area (MAF/A) $m^3/s/km^2$

The 95 percentile flow per unit area (Q95/A) $m^3/s/km^2$

The minimum daily flow per unit area (MDF/A) $m^3/s/km^2$

Long average runoff (LARO) mm

The dependent variables were chosen from three main groups : geomorphic variables, land-cover variables and spectral characteristics. The full set is as follows:

1. The mean of the elevation of the gauging station and the highest point in the catchment (ELEV) in metres above Ordnance Datum.
2. The 10%-85% stream slope (S1085).
3. Stream frequency (STMFRQ) in junctions/km^2.

4-9 inc. The percentage land-cover in six categories estimated from the May 1977 image.

10-15 inc. The same six land-cover categories estimated from the first classification of the April 1984 image.

16-21 inc. The same six land-cover categories estimated from the second classification of the April 1984 image.

22. The average value of reflectance in near infrared of each catchment estimated from the 1977 image (B7).

23. The vegetation index, 1977 bands 5/7 (veg)

24. The biomass index, 1977 bands 7-5/7+5 (bio)

25. The wetness index, 1977 bands 4/6 (wet)

26. The first principal component, 1977 (PC1)

The complete set of data used in the regression analysis is given in Appendix 2.

5.1 REGRESSION ANALYSIS

With only 15 points in each data set it was clearly impossible to include many dependent variables in any model. Thus the first step was to examine inter-relationships between the variables, with a view to reducing the set.

The cross correlation matrix for the 1977 land-cover classification has been given above, in Table 2.7, from which is can be seen that there is only one noticeable correlation in the set, which is the negative correlation between Grass (2) and Arable. The implication that grassland is lost primarily to arable farming is reasonable, but the magnitude of the correlation (-0.71) is not strong enough to warrant leaving out one of these variables. Similarly it was found that there were no strong correlations within the 1984(1) or 1984(2) set of classifications. However, it was felt that some simplification of the classification set was desirable, so on hydrological grounds rather than statistical grounds the two categories Grass (1) and Grass (2) were lumped together to provide an alternative category, Grass.

At this point no choice was made as to which of the three classifications to use. It was decided to make a regression analysis on each land-cover set and see which performed best, before making any selection.

With regard to spectral data, it has been shown above that the data is very highly cross-correlated. The most consistent correlations involve the wetness index, but selection was deferred until it could be seen how each variable performed in regression models.

The correlations of the various independent parameters with the 1977 land-cover set are shown in Table 5.1, and the correlations with the spectral data set in Table 5.2. From these it can be seen that there are few correlations between single parameters that are obviously strong. The exception is the case of Grass (1), which correlates quite strongly with long average runoff, long average rainfall and mean annual flood.

The association here appears to be due to a common link with rainfall i.e. that average runoff and flood flows increase with increasing average rainfall. The classified image was checked and it was found that Grass (1) is confined to the upland areas where rainfall is highest, which probably established the statistical link between grass (1) and rainfall.

All variables were converted to their natural logarithms. The five main independent variables (LARO, a, MAF/A, Q95/A and MDF/A) were then regressed in turn on appropriate groups of the dependent variables i.e. land-cover and spectral properties.

5.2 RESULTS OF THE REGRESSION ANALYSIS

Firstly, the hydrologic variables were regressed in turn on the six land-cover parameters as estimated from the best land-cover classification i.e. the 1977 image. The statistics of the regression are shown in Table 5.3, from which it can be seen that the regressions are mostly good, with the exception of the model for Minimum Daily Flow. Correlation coefficients were greater than 75% (adjusted for degrees of freedom), while standard errors of estimate of the natural logarithms of the dependent variables were small, in the range 0.11 to 0.23.

The regression on the same set of variables was then repeated, using the two land-cover classifications acquired from the 1984 image. The standard errors of estimate given in Table 5.4 show that the quality of modelling is not highly dependent on the image used, and that similar accuracies were obtained from all three land-cover classifications.

An attempt was next made to simplify the model by reducing the set of dependent variables. The strategy here was first to remove 'water' from the set, on the basis that the percentage of open water is very low in all catchments in this region. Secondly, the two grass categories were brought together into one single class, 'grass'. Table 5.5 shows that very little loss in performance was incurred by removing 'water' from the set, except in the model for the losses coefficient, 'a'. However, there was a noticeable drop in performance of the models for LARO, MAF/A and Q95/A when the 'grass' category was used. Similar results were found using the 1984 classifications. The parameter MDF/A had large errors of estimate in all models.

Turning to the spectral variables, it has already been pointed out that all the variables are strongly correlated, and that the most consistent correlation was in the case of the 'wetness' parameter i.e. the band 4/6 ratio. Such strong correlation suggested that the full data set would be redundant, and that it should be possible to achieve maximum performance in the regression model with a much reduced data set. The independent variables were regressed in turn on the spectral data, first using all five variables (wet, veg, bio, b7 and pc1) and then using successively smaller data sets. The variable eliminated at each stage was decided by an analysis of variance, in which the variable making the least contribution to the sums of the squares of the residuals was deleted (i.e. the variable with the smallest F-statistic). Also, in view of the 100% correlation between band 7 and PC1, it was decided to use only band 7, this being the simpler variable to estimate.

Table 5.6 illustrates the sequence of reduction in the data set for each independent variable, and shows the efficiency of each regression as measured by the correlation coefficient and the standard error of estimate. In all cases a reduction in the number of dependent variables can be made, with little loss of model performance (and in some cases a modest improvement in the regression). Generally, a reduction to two parameters can be made with little loss in performance but, unfortunately, the two parameters are not the same in each case. Equally, there is no single parameter which is important in every case.

Taking the regression on all 4 spectral variables as the best model, it was compared with the regressions on land-cover parameters, as in Table 5.7, from which it can be seen that the best regression on spectral variables is considerably worse than the best regression on land-cover variables.

A step-wise regression was then made, forcing the inclusion of the 6 land-cover variables, and entering those spectral variables which gave an improvement in the model. The results are shown in Table 5.8, where it can be seen that the addition of spectral variables gives no significant improvement in the models for 'a', Q95/A or MDF/A. However, in the case of long average runoff and mean annual flood the regression is improved markedly by the addition of 'band 7', 'wet' and 'veg'.

The inclusion of all three gives a regression model with 9 dependent variables which has been based on only 15 data points, which common sense would advise against using. A better approach would be to attempt to improve the land-cover regression by the inclusion of just one spectral parameter. Table 5.9 shows that 'wet' has been included in every stepwise regression, and it has also been shown above that it correlates strongly with every other spectral parameter, so it is clearly a good candidate. On hydrological grounds the inclusion of the biomass index appeared sensible, although statistically it can be seen (Table 5.8) to be of limited value.

These two parameters were added in turn to the set of six land-cover parameters, but the standard error of the regression equations was not changed by any significant amount. (Table 5.9).

The next stage of the analysis was to bring in those terms concerned with the topography of the catchments. In the case of the low-flow and long-average measures, it was felt that the elevation of the catchment was the most important parameter. Thus, the regressions were repeated using the 6-parameter land-cover and 4-parameter spectral models, with the elevation terms added. Somewhat surprisingly, no improvement was found in the regression models. In the case of flood flows, the influence of topography was thought to be much stronger, and so a separate, more thorough investigation was made. (See below).

One of the main hopes of this regression study was to develop models that could be used solely with satellite-derived data. In certain circumstances, though, it would be useful to incorporate rainfall data into the models e.g. the annual runoff model. Extending this idea to average flow statistics, the correlation of rainfall with flow parameters was examined, and it was seen that long average rainfall correlates strongly with long average runoff (93%), fairly strongly with flood flows (77%), but not with low-flows, nor the parameters of the rainfall-runoff model. When rainfall was included in the regression on land-cover parameters a very high correlation was achieved for long average runoff, but the standard error was not improved significantly in any other case (Table 5.10).

5.3 REGRESSION MODELS

Having examined the correlations which exist between individual variables and between groups of variables, the next step was to select the best set of regression models which could be used to predict streamflow.

Where rainfall data is available, then the sequence of annual runoffs from a catchment may be estimated from the series of annual rainfall using the simple model :

ARO =-a + ARF

Where the parameter 'a' may be estimated from a regression on catchment characteristics, using the equations given in Tables 2.13 and 2.14. Similarly, the yield of the catchment (long average runoff may be calculated using the equation

LARO =-a + LARF

Where 'a' has the same value.

Other parameters of streamflow do not benefit significantly from the knowledge of rainfall statistics. They may be estimated from a regression on catchment characteristics alone. Therefore, in the absence of rainfall statistics, streamflow characteristics for this region may be based on a regression on a land-cover classification in which remotely sensed data provides the proportion of area in each catchment within the categories shown below. The model could be simplified by omitting the category 'water' but at the risk of giving poor estimates in the occasional catchment containing large lakes or reservoirs. No consistent improvement can be made in the model by adding spectral characteristics or simple topographic characteristics into the set (except in the case of flood flows which are dealt with later). The coefficients of the regression models (using a transformation to natural logarithms) are given in Table 5.11. It should be noted that the distinction between the two types of natural grassland is an important one, and is necessary for achieving the smallest error of estimate.

Where rainfall data are available, the model for long average runoff is improved by the addition of long average rainfall to the set, which reduces the standard error from 0.11 to 0.06 (on log-transforms). The model then becomes :

$$\ln LARO = -9.69 - 0.0089 \ln water - 0.114 \ln forest - 0.366 \ln g(1) - 0.147 \ln g(2) - 0.0463 \ln crops + 0.008 \ln urban + 2.10 \ln larf$$

$r^2 = 96\%$ s.e.e. = 0.06

Where ground truth is not available, land-use classifications may be erroneous, particularly in regard to the grassland types. In this case it is recommended that a model based simply on spectral characteristics is used. The regression model employing all four of the chosen variables is the most consistent for all streamflow variables. In contrast to the land-cover classification, these variables take values which depend on the particular techniques of image processing that are employed. Therefore the recommendation is that all data should be normalised (i.e. a mean of zero and standard deviation of 1) to bring them to a common standard. Unfortunately this would give negative values, which cannot be transformed to logarithms, therefore it is recommended that an arbitrary scale factor of 5.0 is added to ensure that all values are positive. The coefficients for the normalised and rescaled variables are given in Table 5.12.

Analysis of residuals

No regression model is ever perfect, and it is usually informative to examine the residuals of the model to see where improvements can be made (i.e. the differences between observed values and the values predicted by the regression equation). The residuals of the six regression equations based on land-cover are given in Table 5.13, in which two catchments stand out as having consistently large residuals for each streamflow model. These are the Ebbw and Sirhowy, where the Ebbw catchment shows 7 positive residuals, 5 of them large, and the Sirhowy shows 6 negative residuals, 4 of them large.

This is quite remarkable, since the Sirhowy is a sub-basin of the Ebbw, and would be expected to have very similar characteristics, whereas the model consistently underestimates streamflow for the Ebbw and overestimates it for the Sirhowy. No sensible physical reason could be seen for this, and it was concluded that the pattern of errors occurred by chance. No other consistent feature was seen in the residuals, and so no further modification to the regression models was proposed.

5.4 DISCUSSION OF CORRELATION AND REGRESSION ANALYSIS

The correlation between streamflow characteristics and individual catchment characteristics was generally not strong, with the exception of grass (1) i.e. the molinia grassland, which may be explained by the strong link with rainfall. No strong individual correlations were observed between streamflow statistics and spectral characteristics. Surprisingly, rainfall was strongly correlated only with long average runoff and mean annual flood, and had poor correlation with the measures of low flow.

In a multiple regression analysis on the log-transformed data, the correlations between streamflow and the group of land-cover parameters was generally good. Correlation coefficients were in the order of 80-85%, with the exception of minimum daily flow, which performed poorly in this and all subsequent regressions, possibly due to sampling error or measurement error. It was remarked earlier that the period of record varied widely between stations, and it was shown that the choice of sampling period had a strong effect, on average flows (Chapter 2). The argument is not so strong in the case of minimum flows, because all records included the drought period 1975-76, which will have given the historic lowest flows for all catchments in this area. The problem is more likely to lie in the difficulty of measuring very low flows with gauging structures which are primarily designed to gauge flows in the mid-range. Coupled with this is the fact that very low flows can be radically altered by extraction of water from the river, and the proportion of flow extracted may vary widely from river to river, causing inconsistencies in the flow measurement.

The 95 percentile low-flow is not such an isolated statistic, but is calculated in relation to a range of other flow measurements, and is thus likely to be more stable than the minimum flow. This may be the reason for the large difference in quality of regression for these two measures of low flow. On this basis it is recommended that the Minimum Daily Flow is not used in regression modelling of streamflow.

The regression of streamflow parameters on the group of land-cover parameters remained similar for each of the land-cover classifications made, suggesting that netiher the date of image nor the element of subjectivity involved in making a classification were important factors affecting the quality of the model.

There were very low cross-correlations between the elements of the land-cover set, and so no model was proposed with less than the full set of land-cover parameters. This was logical, in that information from all parts of the catchment would be included in the model. The same is not true, however, of the set of spectral variables, where any one variable contains information from all over the catchment. In this case it was sensible to attempt to reduce the data set, and it was found that the initial set of 5 variables could be reduced to 3 with very little loss of performance of the model (although the best set of 3 was different for each streamflow variable). Since there was such a strong cross-correlation between spectral variables, any one parameter would be useful, and an attempt was made to improve the regression on land-cover parameters by the inclusion of one spectral parameter. In the event, no worthwhile, consistent improvement could be found. Thus the best set of variables remained simply the full set of land cover types.

An interesting feature to emerge from the study was the need to make a distinction between the two types of upland moor vegetation (i.e. Nardus grassland and Molinia grassland). This suggests that certain species or associations could be used as indicators of hydrologic status. Some work has been done in this direction by Clark et al (1986), who have investigated the use of heathland species as surrogates in a hydrological analysis. They have found that the distribution of heathland species can classify combinations of land-slope and soil type, and that this classification can be used to give an indication of Winter Rain Acceptance Potential. It appears that something similar is happening in the South Wales region, where the distribution of Molinia is similarly linked with land-slope and drainage i.e. it colonises the wetter flatter parts of the moorland. Unfortunately, there is a limitation to this approach, which is that candidates for hydrologic surrogates, be they heather or molinia, are not distributed widely over whole regions, and are confined by topography and agriculture to specific locations. Therefore Clark's model can only be successful in the parts of the country where undisturbed heath still exists, and has no application elsewhere.

The same logic applies to the South Wales model, and a better regression model would be one based only on land-cover types of universal application. (The same is also true of some of the American adaptations of the SCS model, in which good results have been achieved by carefully optimising the choice of land-cover types). Thus, in this case, a more generally applicable model would be one based on the five land-cover classes of water, forest, grassland, crops and urban area.

The corollary is that any attempt to make a regression model more general is likely to make the model less precise, and in this case, a reduction to five land-cover parameters gave a significant increase in the standard error of estimate.

At this stage of the work the recommendation must be to use the best model available i.e. the 6-parameter model, but restrict its use to the region of study i.e. South Wales, while at the same time, recognizing the potential for stream-flow models based on a reduced set of universal parameters, which could be developed for use over a more diverse region.

The significance that was found to be attached to the classification of the two grassland types raises another point of discussion. It appears, as was postulated, that the spectral properties of an area may be used to distinguish hydrological features that are otherwise not obvious. In general, though, the spectral regression models did not work very well. This apparent contradiction may be explained by considering the nature of the landscape, which in the area of study has been extensively remodelled by deforestation, afforestation, agricultural development and urbanisation. Thus, the subtle features which should be apparent in the spectral response of an area are overpowered by the coarse variation in present-day land use.

The argument for the use of purely spectral variables depends on a state in which the landscape responds to the hydrologic conditions which prevail, so that what is observed by the satellite is indicative of the hydrology. However, it is now the case in the area of study that man's activities have altered the hydrologic conditions, by field drainage, urban drainage and construction, so that the use to which land is put now determines its hydrological state and not vice versa. The consequence of this is that in the undeveloped areas of the world, where natural vegetation and simple farming practices prevail, spectral models may still be useful, but in the developed areas, the parameters of the streamflow models will be estimated better by land-cover.

In this respect remote sensing still has an important role to play, because efficient techniques exist for the classification of land-cover, while satellite data can provide more timely information than maps. It has been shown that models based on land-cover estimated from satellite imagery provide good errors of estimate, and are robust, in that they are not sensitive to the date or quality of the image. This is fortunate, because classification was shown to be very difficult to carry out with confidence using archived imagery.

The changes in land-cover which are apparent in the satellite imagery highlight a fundamental flaw in this type of modelling. This is due to the assumption made when estimating the 'true' or recorded values on which the regressions are based, which is that the hydrological record is stationary i.e. that the statistical properties of the series are the same at all points in the series. But, if the land use of an area changes, then the hydrological response of the area changes, and the hydrological record is not stationary. Thus, we cannot make a meaningful estimate of such basic variables as the catchment yield or the mean annual flood. The maps of land-use over a period of time that can be interpreted from the imagery give us the opportunity to assess the degree of change that has taken place, and the consequences for the hydrology of the area.

Thus it can be seen that remote sensing can provide useful information for assessing the surface water resources of an area from the point at which a survey just begins in an undeveloped area, up to the time when assessments have to be made of the affect of changing land-use on the hydrology of the area.

The discussion so far has been based on the idea that only data from satellite imagery is available for modelling. The success of the models developed above suggest that the technique would be valuable in parts of the world where mapping and hydrometeorological surveys are inadequate. Clearly, this is not the case in Britain, and it would be foolish to ignore the existence of climatic and topographic data which are easily available.

The use of rainfall data was examined in relation to yield, and simple, reliable models were found for annual runoff and long average runoff, based on rainfall alone. The danger of using an estimate of average runoff based on a simple, fixed function of rainfall is that the relationship may be valid only for the catchments which were used to calibrate it. Any change in the nature of the catchment cannot be reflected by a change in the parameters of the model. The next step, then was to derive a model whose parameters are variable, and change in respect to hydrologically significant features of the catchment. Thus a model was derived in which annual runoff could be calculated on the basis of annual rainfall, in which the parameters were given by a regression model using the catchment characteristics already discussed.

Another approach to using rainfall data was to add it to the set of independent parameters in the regression model. The incorporation of rainfall data into the land-cover model for long average runoff, reduced the standard error to 0.06 on the log transformed data i.e. a 95% confidence limit of about ±12%. However, the addition of rainfall data to the regression models for low flows gave no benefit, which indicates that low flow characteristics in this region may be controlled more by geology than rainfall.

Where topographic map data are available the elevation of the catchment may be calculated, which was thought to be linked to evaporation, and thus influential in determining runoff. However, no significant benefit was achieved by the inclusion of elevation in the regression models. Where the data is in the form of a digital terrain model then opportunities may occur for the combination of slope values and soil characteristics to give assessments of infiltration and Winter Rain Acceptance Potentials which may be beneficial in regression models, but the work was outside the scope of this project. Slope data, however, were found to be useful in flood models.

In conclusion then, this part of the study has shown that spectral characteristics of the catchment, as expressed in satellite imagery are correlated to streamflow and may be used in regression models for estimating streamflow particularly in areas of 'natural' land-cover. Where land-cover has changed, but there is some knowledge of the type of land-cover present, then the spectral characteristics may be used to distinguish 6 hydrologically significant classifications of land-use, which give improved regression models. Where rainfall data are available, then the land-use classifications may be used in a rainfall-runoff model for annual runoff, or alternatively the regression model for long average runoff may be improved.

PART 3 CHAPTER 6 TESTING THE MODELS

The conclusions reached in the preceding section were based on the correlation analyses amongst the various parameters. Having arrived at a choice of preferred models, each model was tested by estimating streamflows for all catchments. The true, or recorded, values were then regressed on the estimated value in each case, to quantify the quality of the estimates. It was recognized that the testing should ideally have been carried out for a completely new set of catchments, but the lack of good quality hydrologic data prevented this from being done.

6.1 FLOOD FLOWS

The success of the regression model for Mean Annual Flood was somewhat surprising when viewed in the light of the generally accepted regression model for Britain i.e. the Flood Studies Report method for estimating flood magnitude from catchment characteristics (NERC, 1975), which includes only one land-cover variable i.e. the proportion of urban land in the catchment. However, some support is given by the American approach as exemplified by the US Soil Conservation Service curve number model (SCS, 1972) which is strongly based on land-cover.

A comparison was thus made of the various available techniques for estimating the Mean Annual Flood (MAF), the 'true' values of which (BESMAF) were determined by calculating the arithmetic mean of the annual maxima as given in the Welsh Water Authority records. These are now 13 years longer than those used in the Flood Studies Report, and give a correspondingly better estimate of the mean.

Firstly the mean annual flood for each catchment (FSRREG) was calculated using the FSR equation for the Wales region, and the parameter values as tabulated in the report. The regression equation is :

$$FSRREG = 0.0213\ AREA^{0.94} STMFRQ^{0.27} S1085^{0.16} SOIL^{1.23} RSMD^{1.03} (1+LAKE)^{-0.85}$$

It was noted that this equation was based on many station records from a variety of regions. To make a fair comparison a local version was calculated, using data from only the 15 study catchments. This was done using the revised BESMAF, but the same values of catchment characteristics as in the Flood Studies Report. The local regression equation was found to be :

$$FSRLOCAL = 0.0059\ AREA^{0.97} STMFRQ^{0.32} S1085^{-0.22} SOIL^{0.16} RSMD^{1.19} (1+LAKE)^{0.76}$$

The change in sign of some of the exponents highlights the dangers of regression modelling on such a small set of data.

The next step was to make an estimate of mean annual flood using the SCS curve number method (SCSMAF). In doing so, the land-cover classifications were based on those found from the image interpretation. The soil classes were based as well as possible on the soil types mapped in the Flood Studies Report. Precipitation volume was found by converting 2 day 5-year return period to 24 hour mean annual rainfall using the data and tables given in Flood Studies Report. A major difficulty lay in estimating a reasonable curve number for each of the Landsat land-cover classes, which were considerably broader than the classes tabulated in the SCS method. Careful selection of data led to a simplified table for curve number (Table 6.1).

Following that analysis, the mean annual floods for each station (LST6 REG) were estimated using the 6 parameter land-cover regression model given in the preceding section. Bearing in mind that topographic data could conveniently be added to this set from a digital terrain model, the regression was repeated using 6 land-cover parameters together with stream frequency and stream slope. The regression equation for the 15 catchments was :

$$LST8REG = 8.17\ water^{0.003} forest^{0.03} g(1)^{0.66} g(2)^{0.58} crops^{0.05} urban^{-0.009}$$
$$stmfrq^{0.27} s1085^{-0.29}$$

However, in view of the very small improvement gained by adding average rainfall to the set no attempt was made to introduce rainfall data.

The values of the mean annual flood estimated using these techniques are given in Table 6.2 and are shown in Fig. 6.1. From this is can be seen that most models work well, with the notable exception of the SCS model, in which all values were greatly under-estimated but in an inconsistent manner. The regional Flood Studies Report performed moderately well, with a correlation between estimated and observed values of 75% (s.e.e. 49.7m³/s). However, when the localised version was used the correlation improved to 94%, with a standard error of only 25m³/s. The regression on 6 Landsat land-cover types performed equally well (94%, 25m³/s) and was only marginally improved by the addition of the two topographic parameters (94%, 24m³/s).

6.2 LOW FLOWS

A similar comparison was made for low flows, this time involving the recorded 95 percentile low flow, the values estimated by the 6-parameter land-cover regression model, and the Low Flow Studies regional equation (Institute of Hydrology, 1979). The latter is a regression model which uses Base Flow Index and average annual rainfall (neither of which can be estimated directly from image interpretation). The estimates are first given in terms of the 10-day 95 percentile flow, with coefficients for this region as follows :

$$Q95(10)/A = (7.6\ BFI^{0.5} + 0.0263SAAR^{0.5} - 2.16)^2$$

A conversion is then made to the 1-day value using :

$$\log_{10} GRADQ95 = 0.023SAAR^{0.5} - 0.19Q95(10)/A^{0.5} - 2.11$$

and $$Q95/A = Q95(10)/A\ (1 - 9.GRADQ95)$$

Values of Base Flow Index and average annual rainfall were extracted from Institute of Hydrology data, and were entered into these equations to give estimates of the unit 95 percentile flow. These estimates are shown in Table 6.3 and Fig. 6.2 together with the estimates made from the 6-parameter land-cover model and the recorded values from Table 3.2.

It can be seen that, whereas the land-cover model provides an adequate estimate (r^2 71%, s.e.e. 1.1m^3/s), the Low Flow Studies model is rather unreliable (s.e.e. 1.53m^3/s). However, when the unit values are scaled up to give estimates of the actual 95 percentile flow the situation is much improved. The land-cover model then has a 98% correlation with recorded values, and provides estimates with a standard error of only 0.14m^3/s. The Low Flow Studies model also gives a good correlation, but its standard error is almost double that of the land-cover model.

6.3 LONG AVERAGE RUNOFF

In the case of yield (long average runoff) there is no generally accepted model against which to compare. The best model suggested by the correlation analysis was one based on the rainfall runoff equation.

$$LARO = -a + LARF$$

Where the parameter 'a' was estimated by a regression on six land-cover classification. This was found to give estimates of long average runoff with a standard error of 57.7mm i.e. a 95% confidence limit of ± 115mm in an average value of 1060mm i.e. 11%. Where rainfall data are not available it must be expected that the accuracy of estimate of yield would fall, and this is confirmed in the model giving yield directly from the 6 land-cover parameters, which has a standard error of 91mm, almost double that of the rainfall-runoff model.

A comparison was made using the proposed 'general' models in which the two types of grassland were combined into one. It would be expected that the benefit of generality is achieved at the cost of a loss of accuracy. However, the model for yield shows a slight improvement in standard error. This perhaps emphasises the point that rainfall is by far the most dominant parameter in this model. In the 95 percentile flow model the expected loss of accuracy is seen.

6.4 MODELS BASED ON SPECTRAL CHARACTERISTICS

Where ground truth is not available the models using spectral data as parameters must be used, but the correlation found in Chapter 5 were considerably weaker than those of the land-cover model. Surprisingly, when their estimates were tested, it was found that the rainfall-runoff model for yield performed best of all those tested, although the model for 95 percentile low flow gave a fairly poor standard error (though still better than the Low Flow Studies estimate).

6.5 CONCLUSIONS

Where rainfall data are available the best estimate of yield is given by a simple runoff relationship :

$$LARO = -a + LARF$$

Where a is the term representing evapotranspiration losses, whose value may be estimated from a regression model on catchment characteristics. Annual values of runoff could be estimated from annual rainfall, using

$$ARO = -a + ARF$$

Where rainfall data are not available, then yield, mean annual flood and the 95 percentile low flow may be estimated directly from a regression on catchment characteristics. The minimum daily flow is not amenable to this treatment.

An efficient set of catchment characteristics may be derived from a small set of hydrologically significant land-cover classes. In the South Wales region the best set is :

Open water, forest, grass(1), grass(2), crops, urban

Where grass (1) represents a vegetation association typical of poorly drained conditions and grass (2) represents a well drained condition. A more general model may be established using

Open water, forest, grassland, crops and urban

but the standard error of estimate increases. The land-cover classes of these models may be derived from interpretations of satellite imagery, in which the classes are identified on the basis of their spectral properties. In cases where insufficient ground-truth exists to make a land-cover classification, then the spectral properties of the catchment may be used directly in the regression models. In which case, yield, mean annual flood and 95 percentile flow may be estimated from a regression on

band 7 , bio , veg , wet

Where band 7 is the mean reflectance in the near infrared and bio, veg and wet are the mean values of band ratios defined above.

The parameter 'a' of the rainfall-runoff relation may also be estimated from similar regression equations, and the same comments apply. The availability of rainfall data, though, does not significantly improve the regression for flood flows or low flows.

The regression model for mean annual flood is only marginally improved by the addition of geomorphic factors such as stream-slope and drainage density.

The performance of the models proposed here is as good as, or better than, well tried and accepted models such as Flood Studies Report model for Mean Annual Flood, the Institute of Hydrology Low Flow Studies model for 95 percentile flow and the USGS SCS Curve Number model for flood flows, even when these models are modified for the same data as the regression models.

River	Period of record	Equivalent sampling error in Ebbw record %
Honddu	1964-81	-4.9
Usk	1966-78	-8.6
Lwyd	1967-84	-1.1
Usk	1964-73	-6.4
Sirhowy	1973-81	-8.2
Grwyne	1971-81	-9.0
Yscir	1973-84	-0.7
Cynon	1963-84	-1.3
Taff	1971-81	-2.3
Ogmore	1964-84	-9.7
Mellte	1972-85	-0.4
Dulais	1972-85	-0.4
Ewenny	1972-84	-0.4

TABLE 3.1 Equivalent sampling errors of records shorter than 1958-84

Station Name	Station Number	Catchment area (km^2)	Period of record (low flows)	Period of record (flood flows)	Mean Elevation (m)	LARO (mm)	LARF (mm)	MAF (m^3/s)	ADF (m^3/s)	Q95 (m^3/s)	MDF (m^3/s)
Ebbw at Rhiwderin	56002	216.5	1958-84	1958-79	320	1035	1483	106.2	8.030	1.512	1.046
Honddu at Brecon	56003	62.1	1964-81	1964-80	309	749	1153	24.6	1.480	0.159	0.045
Usk at Llandetty	56004	543.9	1966-78	1966-79	495	970	1485	322.8	16.721	2.358	0.366
Lwyd at Ponthir	56005	98.1	1967-84	1967-83	298	986	1415	52.2	3.102	0.657	0.339
Usk at Trallong	56006	183.8	1964-78	1964-81	480	1179	1666	167.0	6.558	1.001	0.341
Sirhowy at Wattsville	56011	76.1	1973-81	1971-80	340	857	1454	41.0	2.076	0.341	0.105
Grwyne at Millbrook	56012	82.2	1971-81	1971-83	447	772	1251	23.0	2.034	0.344	0.152
Yscir at Pont-ar-Yscir	56013	62.8	1973-84	1973-85	318	951	1427	40.5	1.921	0.183	0.075
Cynon at Abercynon	57004	106.0	1963-84	1960-84	301	1189	1766	81.3	4.082	0.561	0.283
Taff at Pontypridd	57005	454.8	1971-84	1969-83	466	1245	1822	286.3	18.368	3.546	1.699
Ogmore at Bridgend	58001	158.0	1964-84	1960-82	291	1238	1739	108.5	5.85	1.504	0.471
Neath at Resolven	58002	190.9	-	1960-82	375	NR	NR	182.8	9.213	1.605	0.278
Mellte at Pont-neath-vaughn	58006	65.8	1972-85	1971-86	382	1433	2061	72.1	2.831	0.340	0.196
Dulais at Cilfrew	58008	43.0	1972-85	1972-86	262	1356	1759	43.1	1.759	0.255	0.127
Ewenny at Keepers Lodge	58009	62.5	1972-84	1960-83	154	869	1328	23.7	1.583	0.320	0.163
Coefficient of variation					0.28	0.20	0.16	0.91	0.95	0.99	1.16

NOTES ON TABLE 3.2

Station number : The reference number for the gauging station used in the national archive of surface water data (Institute of Hydrology).

LARO : Long average run-off i.e. average annual outflow from the catchment, over the period of record, expressed in mm over the catchment.

LARF : Long average rainfall i.e. the average annual precipitation over the catchment, over the period of record, expressed in mm over the catchment.

Mean elevation: average of the elevation of the gauging station and the level of the highest point in the catchment area.

MAF : Mean annual flood i.e. the arithmetic mean of the series of maximum instantaneous discharges in each calendar year of record.

ADF : Average daily flow i.e. the average of all available daily mean flows over the period of record.

Q95 : The 95 percentile flow i.e. the discharge which was equalled or exceeded for 95% of the period of record.

MDF : Minimum daily flow i.e. the minimum daily mean flow in the period of record.

TABLE 3.2 Summary hydrological data

YEAR	EBBW RF	EBBW RO	HONDDU RF	HONDDU RO	USK (LLANDETTY) RF	USK (LLANDETTY) RO	LWYD RF	LWYD RO	USK (TRALLONG) RF	USK (TRALLONG) RO	SIRHOWY RF	SIRHOWY RO	GRWYNE RF	GRWYNE RO	YSCIR RF	YSCIR RO
1958	1574	1080														
1959	1539	1103														
1960	1945	1408														
1961	1429	988														
1962	1257	795														
1963	1451	964														
1964	1165	674	885	447					1667	1276						
1965	1585	1178	1209	869					1713	1349						
1966	1608	1255	1361	937	1301	814			1505	1119						
1967	1688	1319	1345	927	1771	1369	1608	1090	1336	934						
1968	1580	1045	1154	747	1887	1521	1535	1066	1517	1095						
1969	1247	783	1093	727	1995	1578	1175	799	1286	797						
1970	1473	1136	1222	818	1704	1223	1402	965	1711	1247						
1971	1277	752	996	596	1505	1012	1256	778	1149	660			1137	693		
1972	1611	1069	1210	882	1792	1241	1596	1106	1925	1476			1470	912		
1973	984	511	909	551	1489	853	977	513	1251	773	1052	459	853	431	1122	646
1974	1634	1065	1450	1048	1941	1428	1604	1074			1685	1078	1447	910	1784	1239
1975	1076	595	868	580	1311	779	1071	646			1118	615	993	551	1137	679
1976	1318	847	957	515	2210	1768	1317	938			1344	708	1260	703	1193	651
1977	1504	1069	1215	754	1418	948	1488	1101			1511	886	1415	896	1482	1002
1978	1334	896	1020	601	1339	776	1274	898			1422	787	1139	683	1225	817
1979	1476	1123	1338	945			1476	1094			1603	1020	1358	963	1548	1144
1980	1662	1234	1252	752			1511	1264			1653	1069	1356	913	1455	998
1981	1696	1243	1296	786			1575	1195			1696	1094	1328	832	1544	1097
1982	1858	1540					1749	1267							1704	1156
1983	1571	1222					1440	1006							1460	1029
1984	1495	1039					1408	954							1470	958

TABLE 3.3 ANNUAL RAINFALL AND RUNOFF (mm)

YEAR	CYNON RF	CYNON RO	TAFF RF	TAFF RO	OGMORE RF	OGMORE RO	MELLTE RF	MELLTE RO	DULAIS RF	DULAIS RO	EWENNY RF	EWENNY RO
1963	1666	1015										
1964	1504	759			1458	988						
1965	1976	983			1943	1505						
1966	2021	1309			1906	1424						
1967	2181	1443			2097	1643						
1968	1517	1174			1719	1183						
1969	1577	944			1541	1000						
1970	1927	1311			2030	1476						
1971	1542	967	1587	997	1560	1001						
1972	1949	1462	2013	1409	1836	1261	2164	1518	1863	1488	1413	797
1973	1282	644	1301	711	1303	792	1478	951	1314	966	1040	523
1974	2217	1635	2268	1568	2038	1539	2476	1828	2051	1623	1569	932
1975	1373	861	1414	896	1385	998	1662	1012	1397	1030	1044	600
1976	1439	886	1464	871	1460	840	1509	988	1312	906	1227	613
1977	1757	1228	1877	1270	1660	1141	2055	1521	1865	1422	1282	784
1978	1542	1029	1659	1099	1592	1049	1827	1263	1555	1138	1271	794
1979	1824	1342	1901	1406	1862	1372	2064	1574	1863	1534	1444	1005
1980	1843	1360	1968	1387	1887	1351	2249	1604	1877	1444	1431	1039
1981	1952	1484	2087	1549	2016	1573	2347	1679	1992	1636	1505	1179
1982	2094	1668	2212	1620	1865	1403	2497	1650	2105	1613	1374	1050
1983	1953	1383	1994	1384	1708	1240	2261	1565	1837	1492	1347	1018
1984	1722	1263	1764	1261	1662	1226	1977	1319	1677	1215	1321	957
1985							2292	1590	1911	1476		

TABLE 3.3 ANNUAL RAINFALL AND RUNOFF (mm)

	A	B	r^2	s.e.e.	a	s.e.e
Ebbw	532	1.06	92.6	67.8	448.3	67.7
Honddu	293	0.90	90.1	55.3	405.4	56.4
Usk (Llandetty)	551	1.02	95.7	50.0	487.9	48.4
Lwyd	344	0.94	89.1	68.8	428.2	67.8
Usk (Trallong)	693	1.12	96.5	66.2	514.7	73.0
Sirhowy	508	0.94	96.9	40.3	596.4	40.4
Grwyne	277	0.84	91.8	49.6	479.0	56.8
Yscir	376	0.93	93.8	51.5	475.7	51.3
Cynon	440	0.92	76.6	140.0	577.6	138.0
Taff	464	0.94	97.0	51.4	577.2	52.7
Ogmore	560	1.03	95.0	56.9	501.1	56.1
Neath (no records)						
Mellte	240	0.81	93.2	73.2	628.3	94.3
Dulais	300	0.94	96.3	49.0	402.6	49.5
Ewenny	528	1.05	68.8	117.0	459.8	112.7

TABLE 3.4 Linear regression analysis of annual runoff on annual rainfall

Station	r(1)	95% C.L.
Ebbw	-0.005	± 0.091
Honddu	-0.036	± 0.112
Lwyd	-0.027	± 0.112
Cynon	0.032	± 0.101
Ogmore	0.068	± 0.103

TABLE 3.5 Lag-one serial correlation coefficient r(1) and 95% confidence limits

Catchment N°.	WATER	FOREST	GRASS(1)	GRASS(2)	ARABLE	URBAN
1	0.17	25.88	10.69	42.90	8.21	12.14
2	0.00	6.93	60.51	25.77	6.79	0.00
3	0.39	12.03	42.51	34.65	10.16	0.26
4	0.00	16.43	15.85	39.09	12.15	16.47
5	0.57	13.05	38.05	33.79	14.38	0.16
6	0.03	18.56	20.97	38.22	11.54	10.69
7	0.03	31.84	21.36	41.90	4.88	0.00
8	0.00	5.59	34.45	47.46	12.17	0.04
9	0.02	25.81	8.81	38.00	16.63	10.73
10	0.43	20.65	10.35	40.01	18.16	10.40
11	0.00	21.38	28.76	26.89	15.68	7.29
12	0.15	27.75	5.01	39.40	21.55	6.14
13	0.20	6.74	6.38	58.47	27.95	0.26
14	0.09	24.82	8.53	41.33	14.01	11.22
15	0.00	7.10	55.28	16.13	15.36	6.12

TABLE 4.1 Landcover classification (%) 27/5/77

	WATER	FOREST	GRASS(1)	GRASS(2)	ARABLE
Forest	-0.062				
Grass(1)	0.213	-0.152			
Grass(2)	0.138	0.131	0.338		
Arable	-0.047	-0.643	-0.447	-0.705	
Urban	-0.241	0.405	0.016	-0.062	-0.475
Variance over 15 catchments	3.2	75.5	34.1	96.4	324.0

TABLE 4.2 Cross-correlation matrix of land classification 27/5/77

	Wet	Veg	Bio	Band 7
veg	0.958			
bio	-0.947	-0.998		
band 7	-0.929	-0.868	0.866	
pcl	-0.923	-0.860	0.858	1.000

TABLE 4.3 Cross-correlation of spectral data, 1977 image

	LARO	LARF	a	MAF/A	Q95/A	MDF/A
Water	0.370	0.392	0.497	0.349	0.152	0.025
Forest	0.106	0.089	-0.006	0.015	0.510	0.340
Grass(1)	0.793	0.860	-0.060	0.768	0.419	0.214
Grass(2)	0.432	0.488	-0.471	0.475	-0.053	0.192
Crops	-0.615	-0.646	0.262	-0.513	-0.523	-0.545
Urban	0.151	0.123	0.102	-0.086	0.524	0.677
LARF	0.964	1.000	0.037	0.862	0.515	0.436

TABLE 5.1 Cross-correlation of parameters

	LARO	LARF	a	MAF/A	Q95/A	MDF/A
Wet	0.651	0.660	-0.103	0.509	0.639	0.655
Veg	0.700	0.742	-0.232	0.594	0.533	0.568
Bio	-0.695	-0.738	0.246	-0.594	-0.505	-0.564
B7	-0.468	-0.486	0.161	-0.323	-0.595	-0.631
PC1	-0.462	-0.479	0.163	-0.318	-0.595	-0.630

TABLE 5.2 Cross-correlations of parameters

	Correlation coeff r^2	Standard error of estimate
LARO	84	0.11
a	85	0.17
MAF/A	81	0.23
Q95/A	79	0.22
MDF/A	49	0.53

TABLE 5.3 Results of regression on the 6-parameter 1977 land-cover classification

	LARO	a	MAF/A	Q95/A	MDF/A
1977	0.11	0.17	0.23	0.22	0.55
1984(1)	0.11	0.16	0.22	0.24	0.55
1984(2)	0.13	0.17	0.28	0.24	0.51

TABLE 5.4 Standard error of estimate of 6-parameter regressions

		Standard error of estimate				
		LARO	a	MAF/A	Q95/A	MDF/A
1977	Water,forest,g(1),g(2),crop,urban	0.11	0.17	0.23	0.22	0.55
1977	forest,g(1),g(2),crop,urban	0.11	0.21	0.22	0.21	0.53
1977	Water,forest,grass,crop,urban	0.16	0.16	0.31	0.25	0.53
1977	forest,grass,crop,urban	0.15	0.20	0.30	0.24	0.52
1984(1)	Water,forest,g(1),g(2),crop,urban	0.11	0.16	0.22	0.24	0.53
1984(1)	forest,grass,crop,urban	0.15	0.22	0.28	0.24	0.49

TABLE 5.5 Simplifications of sets of dependent variables

Independent variable	Wet	Veg	Bio	B7	r^2	s.e.e
LARO	0.33	0.25	0.19	1.48	62	0.15
	1.16	0.77		1.83	62	0.14
	3.25			2.08	60	0.14
	3.10				44	0.16
a	-0.73	1.84	2.05	0.34	55	0.26
	-1.35	2.35	2.57		55	0.25
		2.86	2.93		46	0.26
			0.90		6	0.32
MAF/A	-0.08	0.87	0.74	1.94	62	0.29
		1.68	1.23	2.66	62	0.28
		3.58		2.31	57	0.29
		2.78			37	0.33
Q95/A	-0.07	1.53	1.74	-0.13	69	0.24
		3.07	3.05	-0.07	69	0.23
		3.87	3.61		69	0.22
		2.72			36	0.30
MDF/A	1.49	-0.99	-0.89	0.27	55	0.46
	1.76	-0.99	-0.89		55	0.44
	1.90	-0.97			52	0.44
	3.47				48	0.44

TABLE 5.6 t-values of the stepwise regression

	4-parameter spectral		6-parameter land-cover	
	r^2	s.e.e.	r^2	s.s.e.
LARO	62	0.15	84	0.11
a	55	0.26	85	0.17
b	60	0.07	77	0.06
MAF/A	62	0.24	81	0.23
Q95/A	69	0.24	79	0.22
MDF/A	55	0.46	49	0.55

TABLE 5.7 Comparison of regression models

	Wet	Veg	Bio	B7	r^2	s.e.e
LARO					84	.11
				*	92	.08
	*			*	94	.08
	*	*		*	94	.09
a					85	.17
	*				86	.18
	*		*		89	.17
MAF/A					81	.23
				*	89	.19
	*			*	94	.15
	*	*		*	98	.10
Q95/A					79	.22
			*		79	.23
	*		*		80	.25
MDF/A					79	.22
			*		79	.23
	*		*		80	.25

TABLE 5.8 Supplementing the land-cover set

	s.e.e. 6-parameter land-cover	6-parameter + Biomass	6-paramter + wet
LARO	0.11	0.10	0.11
a	0.17	0.18	0.18
MAF/A	0.23	0.21	0.24
Q95/A	0.22	0.23	0.24
MDF/A	0.55	0.56	0.55

TABLE 5.9 Effect of supplementing the land-cover data set

	LARF		6 land-cover		6 land-cover + LARF	
LARO	93	.06	84	.11	96	.06
a	-		86	.17	86	.18
MAF/A	77	.20	81	.23	85	.22
Q95/A	35	.30	79	.22	80	.23
MDF/A	26	.54	49	.55	72	.47

TABLE 5.10 Correlations coefficients (%) and standard error for regressions on natural log transforms

	Constant	Coefficient					
		Water	Forest	Grass(1)	Grass(2)	Crops	Urban
LARO	8.26	0.0048	0.133	0.448	0.110	0.0002	-0.007
a	8.57	0.036	0.223	0.174	0.208	0.567	0.541
MAF/A	2.82	0.0118	0.194	0.913	0.627	0.212	-0.007
Q95/A	3.379	-0.0061	0.435	0.430	-0.114	0.043	0.021
MDF/A	-0.581	-0.0297	0.006	-0.242	-0.193	-0.410	0.071

Example ln LARO = 8.26+0.005 ln Water +0.13 ln forest +0.45 lng(1) +0.11 ln g(2) +0.0002 ln crops -0.007 ln urban

TABLE 5.11 Regression models - landcover

	Constant	Coefficient			
		Wet^1	Veg^1	Bio^1	$B7^1$
LARO	2.369	1.000	0.701	0.098	1.085
a	-17.39	1.770	5.213	6.206	1.538
MAF/A	-19.83	0.205	6.176	3.116	2.663
Q95/A	-5.99	3.354	-0.452	1.345	0.571
MDF/A	6.96	5.647	-6.643	-3.165	0.257

Example $\ln LARO = 2.369 + 1.00 \ln wet^1 + 0.70 \ln veg^1 + 0.10$ lb $Bio^1 + 1.09 \ln B7^1$

where wet^1 = normalised wet + 5.00

TABLE 5.12 Regression models - spectral

Catchment	LARO	a	MAF/A	Q95/A	MDF/A
1	1.28753	2.14092	0.08117	1.59882	1.18363
2	0.50830	0.16029	1.00195	-0.39789	-0.88845
3	0.31171	-1.08130	0.18291	-0.00404	-1.54588
4	-0.25840	-0.94745	-0.23979	0.80993	0.22676
5	0.65612	0.86681	0.94211	0.37931	1.12574
6	-1.85496	-0.51798	-0.58629	-1.34453	-1.04850
7	-1.18009	-0.48383	-1.63059	0.39183	1.21124
8	0.28089	0.83154	0.18191	-0.77640	-0.49799
9	-0.64456	1.00729	-0.02409	-1.83665	-0.29070
10	-0.46670	0.18873	-1.48371	0.50937	0.77368
11	1.30725	0.51842	0.24338	1.55376	0.70440
12	*	*	0.36811	0.06290	-1.90116
13	-0.27508	-0.23986	-0.68615	1.07644	1.26926
14	1.34331	-1.63846	2.07285	-0.68319	-0.05447
15	-0.93817	-0.19408	-1.21437	-0.00274	1.04853

TABLE 5.13 Standardized residuals of the land-cover regression equations

	Soil Group A	B	C	D
Forest	35	60	73	80
Grass (1)	60	74	79	86
Grass (2)	44	65	77	82
Crops	63	74	82	85
Urban	82	88	91	93

TABLE 6.1 SCS curve numbers

Note: In this table the soil groups follow the classification :

Soil Group	Min Infiltration Rate	Description
A	0.30 - 0.45	Aggregated silts
B	0.15 - 0.30	Sandy loam
C	0.05 - 0.15	Clay loams
D	0.00 - 0.05	Heavy plastic clays

Catchment	BESMAF	FRSREG	SCSMAF	LST6REG	LST8REG	FSRLOCAL
Ebbw	105.2	125.6	37.9	97.8	92.7	95.3
Honddu	24.6	24.2	16.8	18.9	20.1	26.6
Usk (Llandetty)	322.8	203.4	68.5	312.7	340.7	325.3
Lwyd	52.2	55.3	10.8	54.4	48.3	48.4
Usk (Trallong)	167.0	74.0	8.7	142.9	131.3	128.5
Sirhowy	41.0	48.1	28.8	46.3	44.5	43.1
Grwyne	23.0	42.2	12.2	28.9	26.9	39.7
Yscir	40.5	31.8	17.8	39.4	43.0	31.0
Cynon	81.3	115.5	56.9	79.2	90.8	92.5
Taff	286.3	373.3	133.4	390.8	393.4	372.3
Ogmore	108.5	117.9	126.7	105.1	103.8	90.1
Neath	182.8	218.6	99.7	107.9	168.6	157.3
Mellte	72.1	77.9	76.5	78.5	72.0	51.0
Dulais	45.1	60.8	29.0	29.0	33.8	52.7
Ewenny	23.7	20.8	29.9	27.5	26.3	23.1
r^2		75	37	94	94	94
s.e.e.		49.7	78.3	24.6	23.6	24.9

TABLE 6.2 Estimates of mean annual flood (m^3/s)

Parameter being estimated	Model	Correlation coeff. r^2	s.e.e.
Mean Annual Flood (MAF)	6-parameter land-cover	94.0	24.6 m^3/s
	FSR local version	94.0	24.9 m^3/s
Long-average runoff (LARO)	Rainfall-runoff 6 land-cover	94.7	57.7 mm
	Rainfall-runoff 5 land-cover	94.7	52.0 mm
	Rainfall-runoff 4 spectral	96.0	45.1 mm
	6-parameter land-cover	83.6	90.9
95 percentile low-flow per unit area (Q95/A)	6-parameter land-cover	71.0	1.08 m^3/s
	5-parameter land-cover	57.9	1.31 m^3/s
	4-parameter spectral	60.0	1.29 m^3/s
	Low Flow Studies	42.4	1.53 m^3/s
95 percentile low-flow (Q95)	6-parameter land-cover	98.1	0.14 m^3/s
	Low Flow Studies	92.7	0.27 m^3/s

TABLE 6.3 Correlation between recorded and estimated values of streamflow

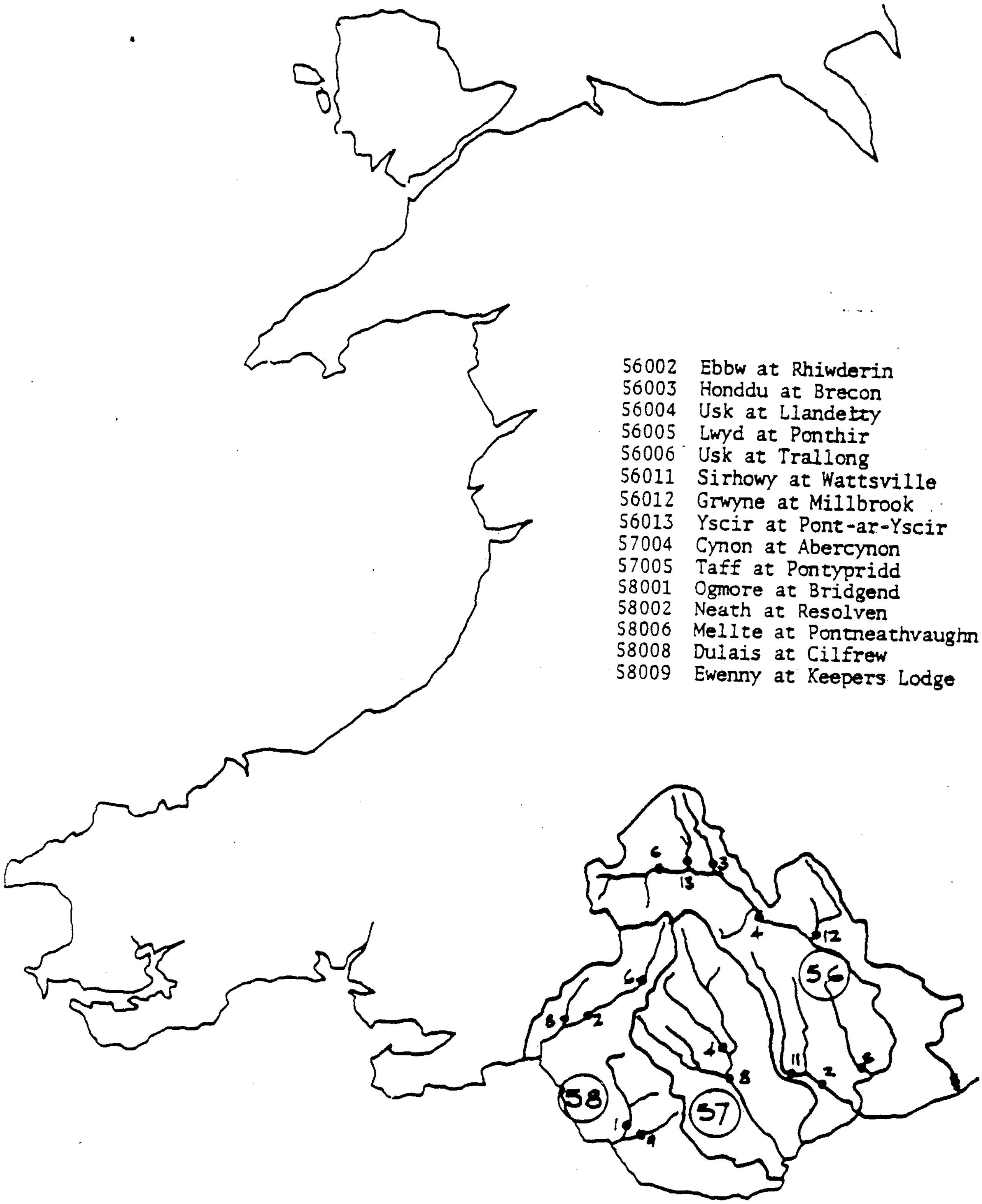


Fig. 3.1 STUDY CATCHMENTS

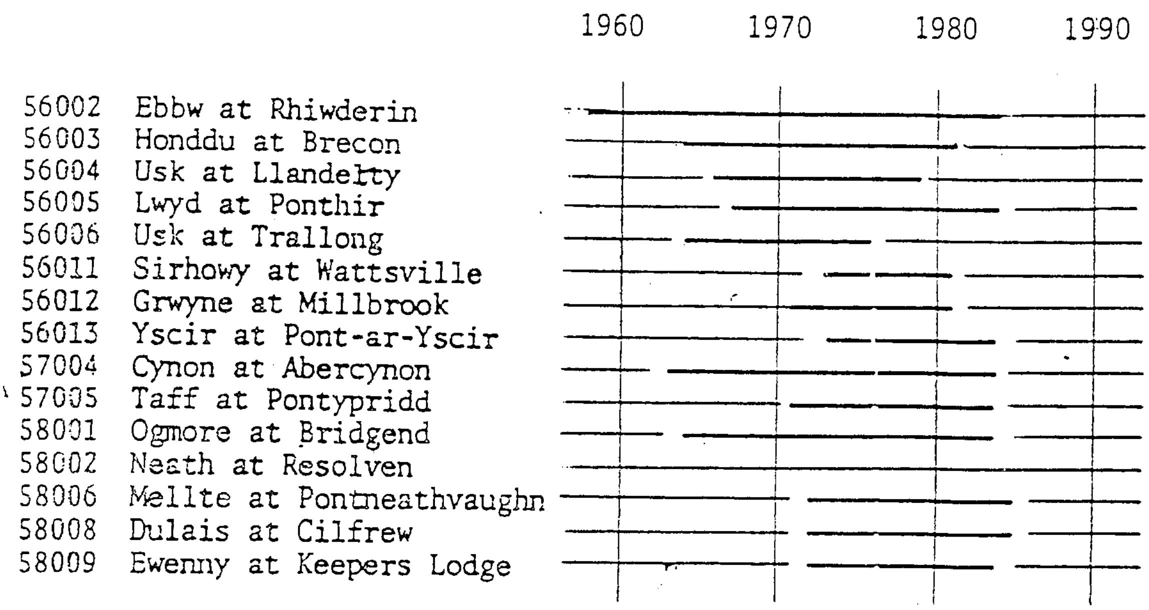


Fig. 3.2 Periods of record

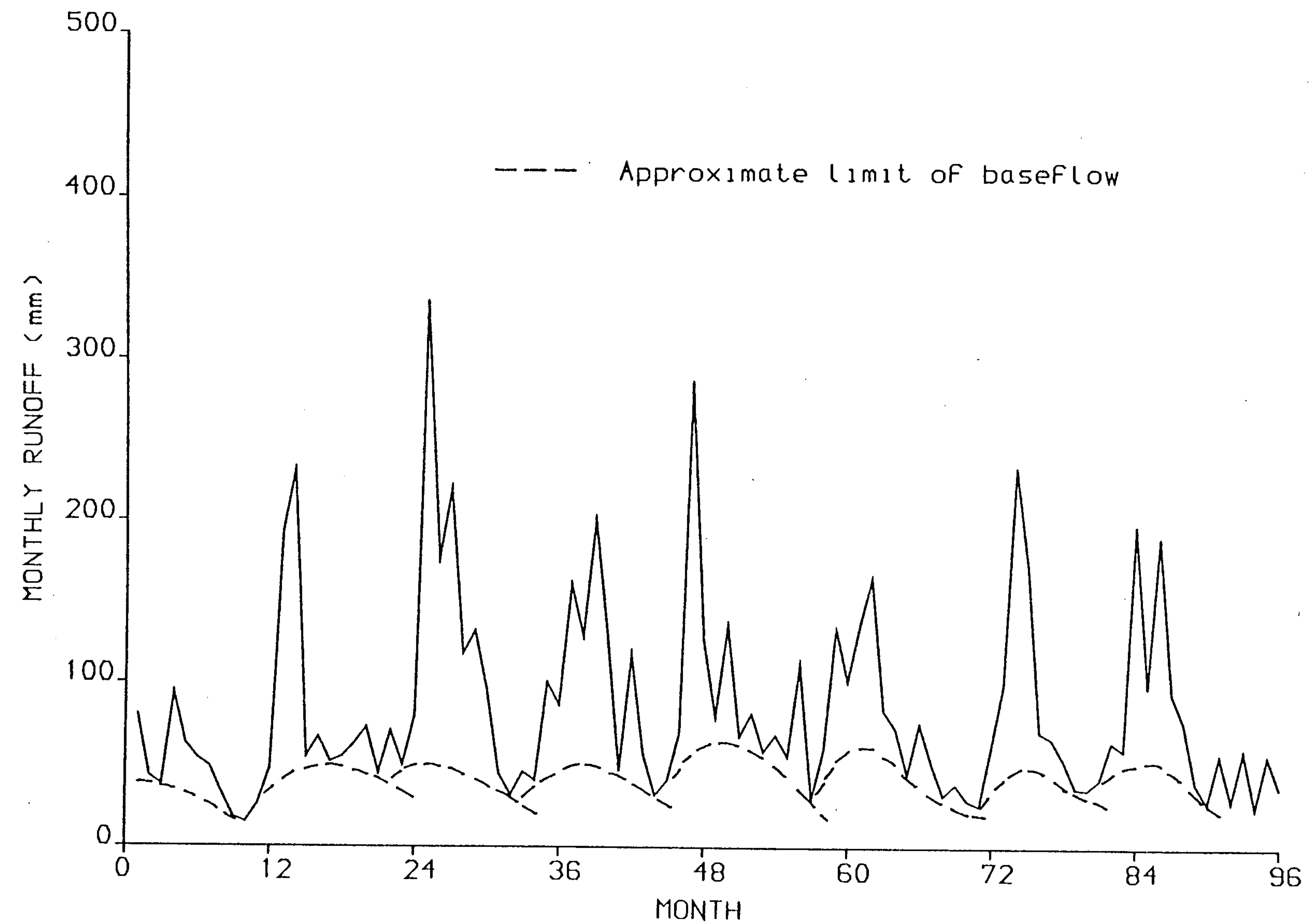


FIG. 3.3 EBBW RIVER MONTHLY RUNOFF 1964-1972

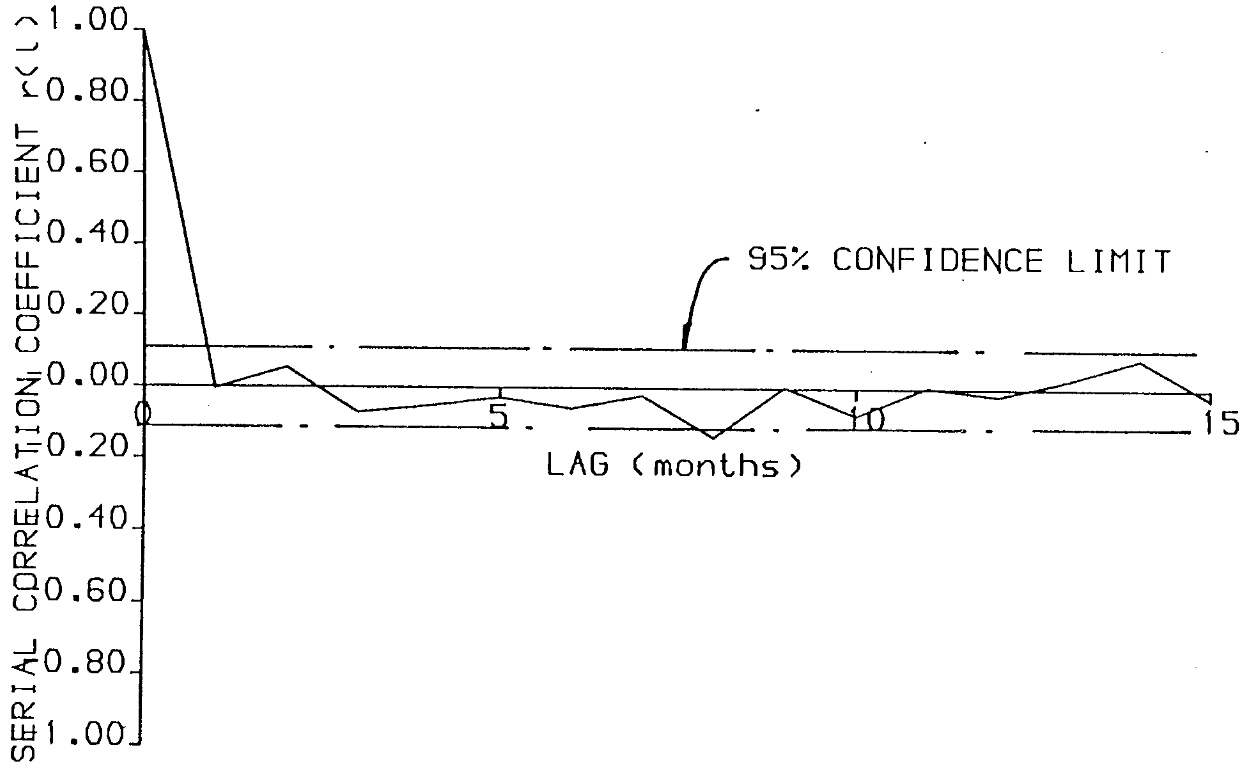


FIG. 3.4 SERIAL CORRELOGRAM EBBW RIVER

70

Aston University

Content has been removed for copyright reasons

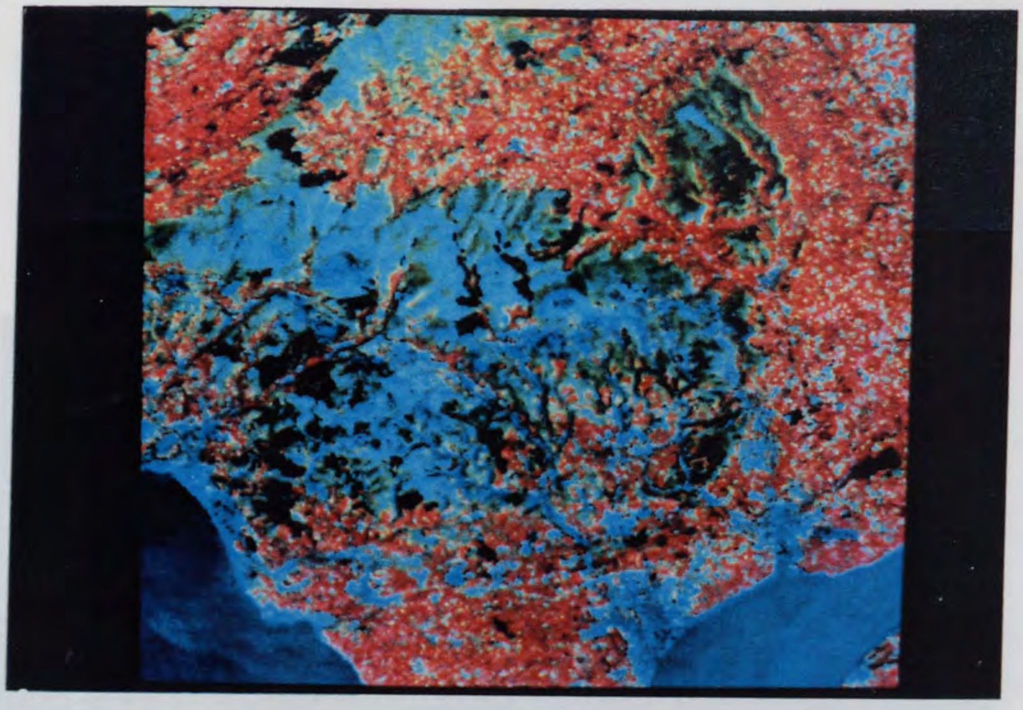
Aston University

Content has been removed for copyright reasons

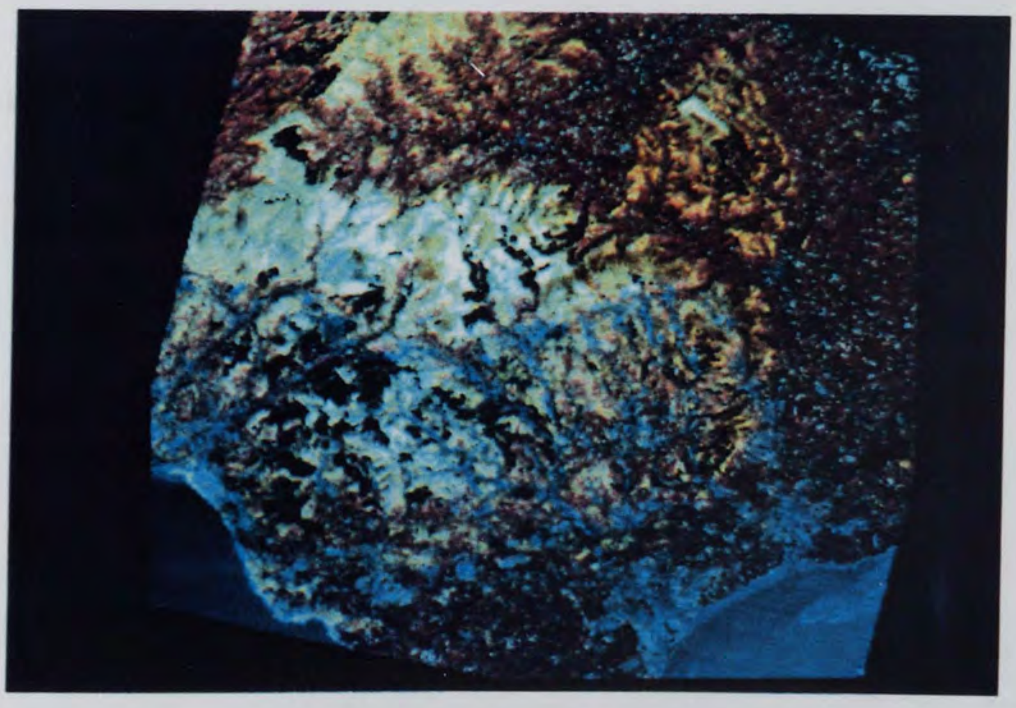
(nm)

Fig. 4.1 Spectral characteristics of soils

(from Barrett, 1982)



(a) False Colour Composite May 1977



(b) False Colour Composite April 1984

Fig. 4.2 South Wales Imagery

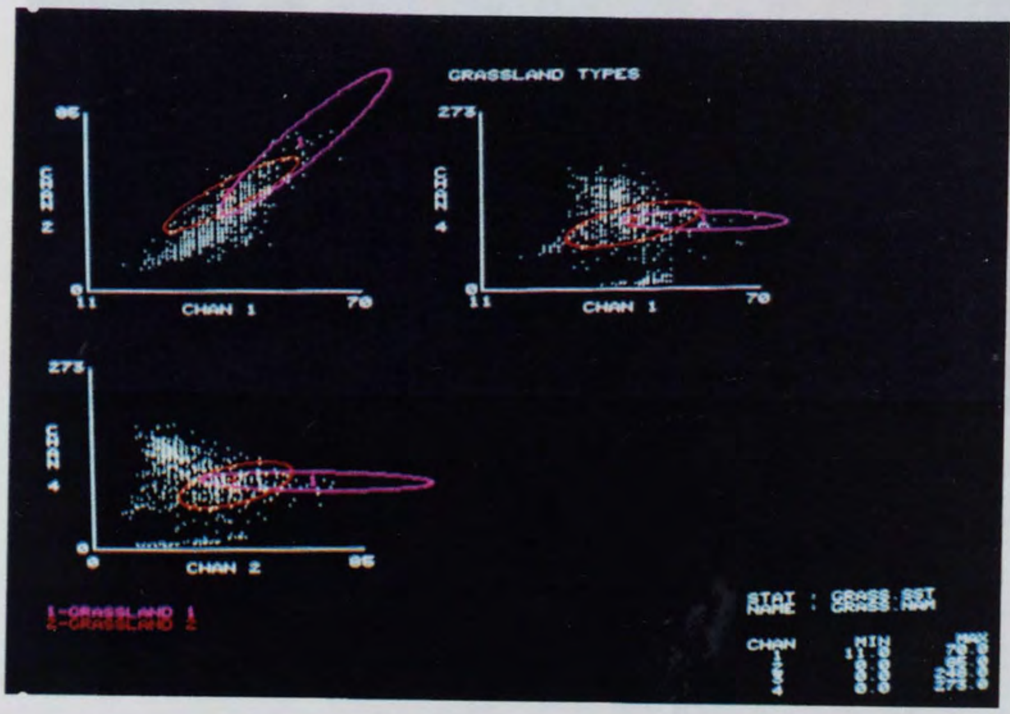


Fig. 4.3 Feature space of May 1977 image.
(Distinction of two grassland types)

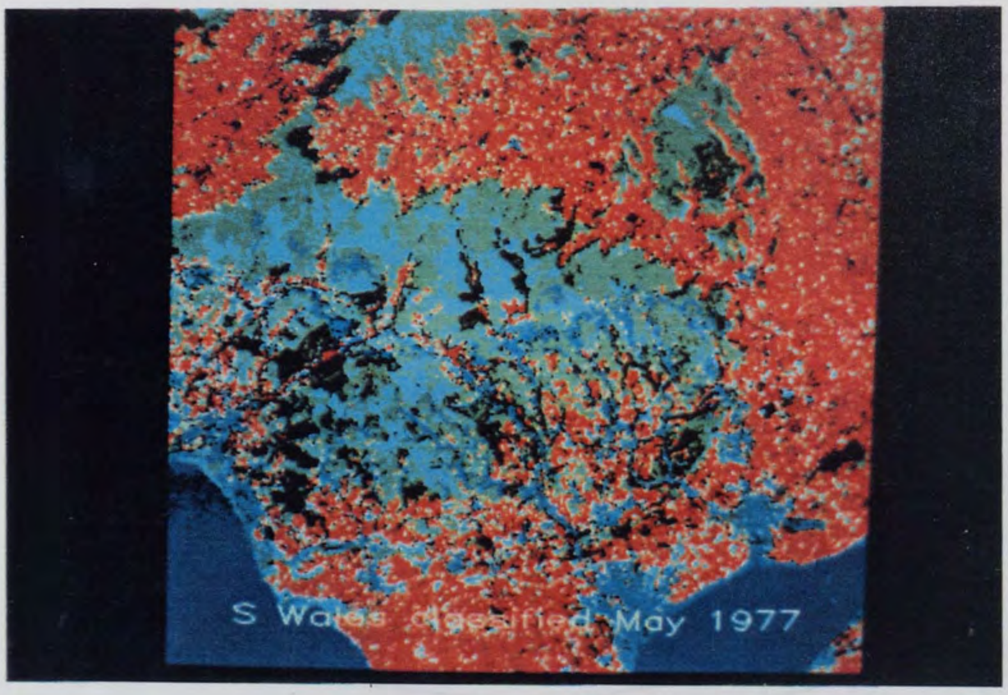


Fig. 4.4 Land-use classification South Wales (1977)

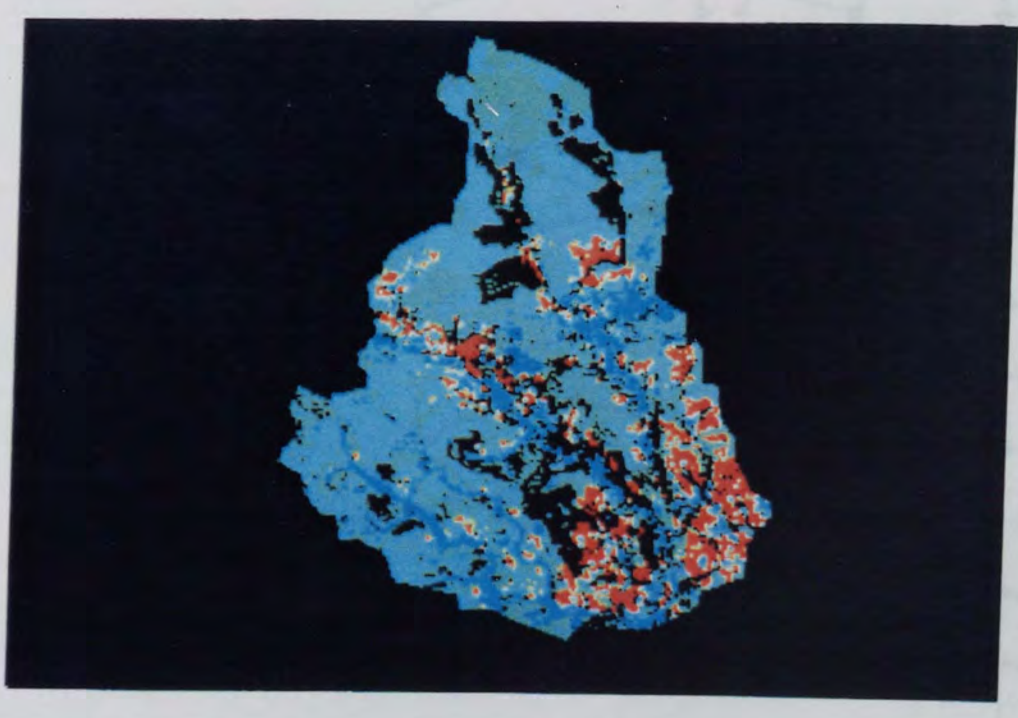


Fig 4.5 Land-use classification Taff basin (1977)

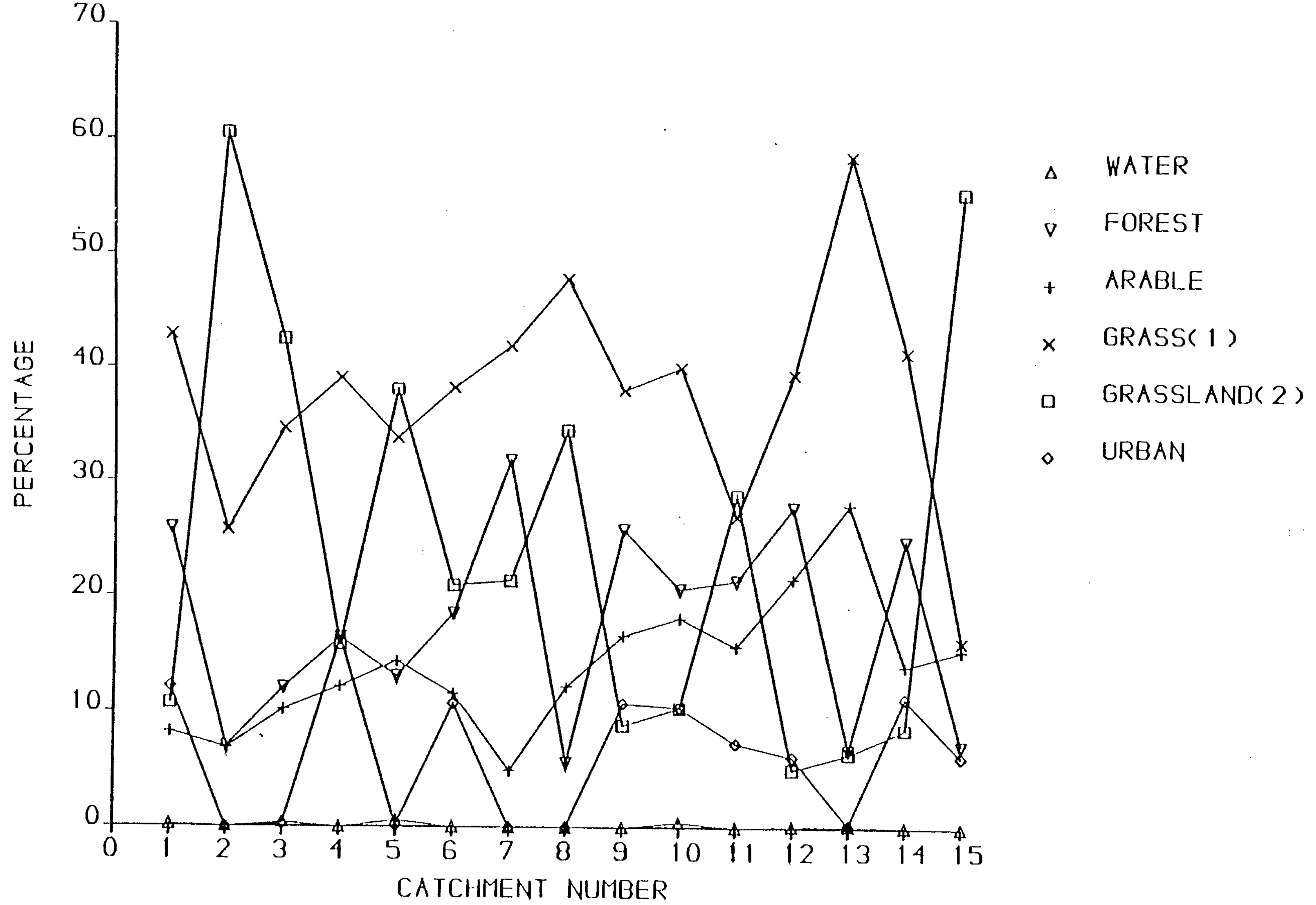


FIG. 4.6 PERCENTAGES OF LAND COVER (CLASSIFICATION OF 1977 IMAGE)

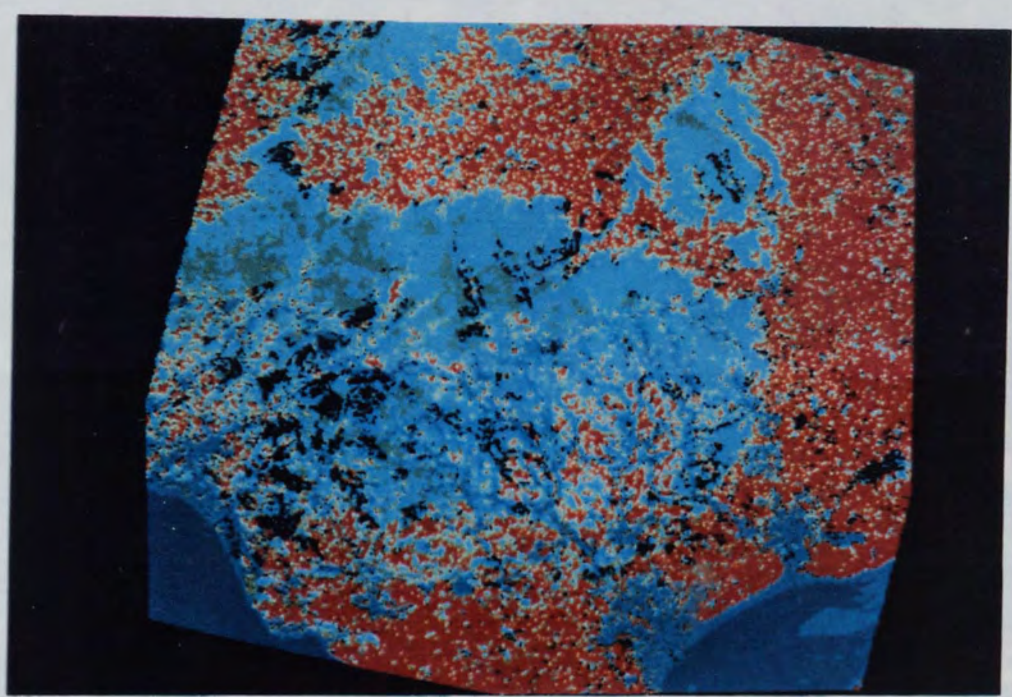
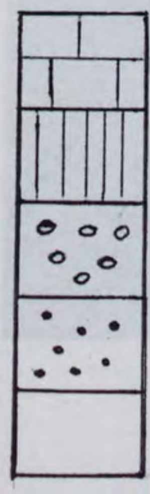


Fig. 4.7 Land-use classification South Wales (1984)

Fig. 4.8 Brecon Beacons: Geology and Transects

KEY



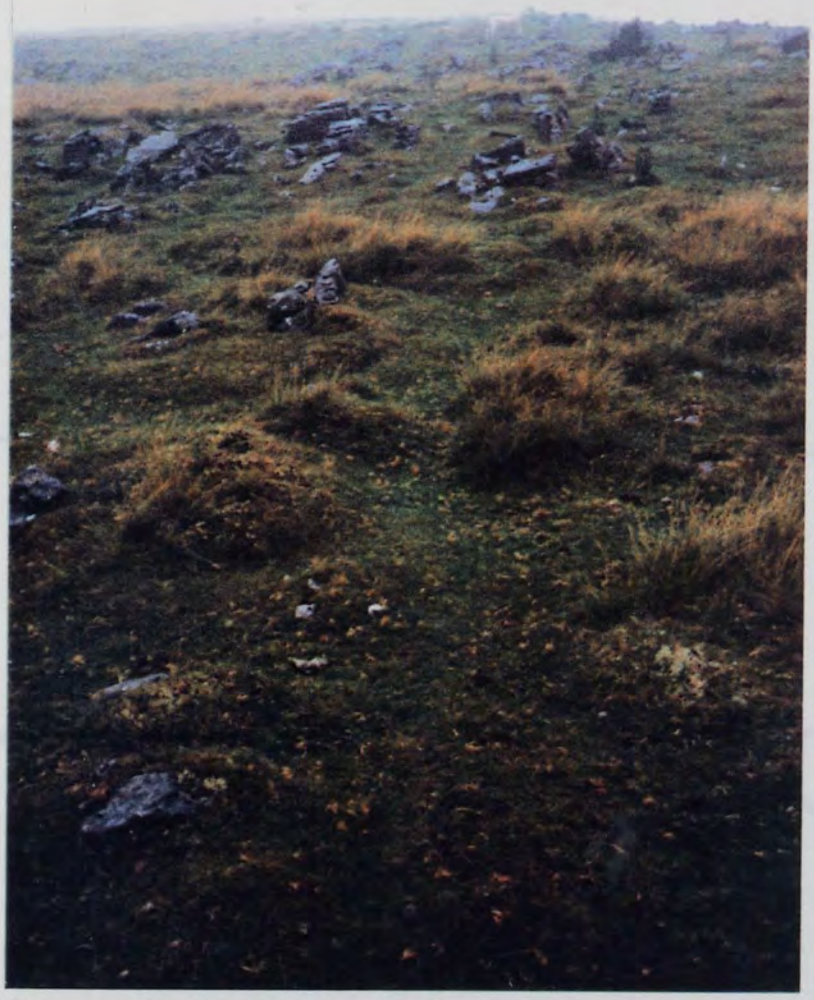
Carboniferous Limestone

Limestone-Shales

Grey Grits

Plateau Beds

Brownstones



Transect 1: Nardus-Fescue

Transect 2: Molinia-Sphagnum

Fig. 4.9 Vegetation types

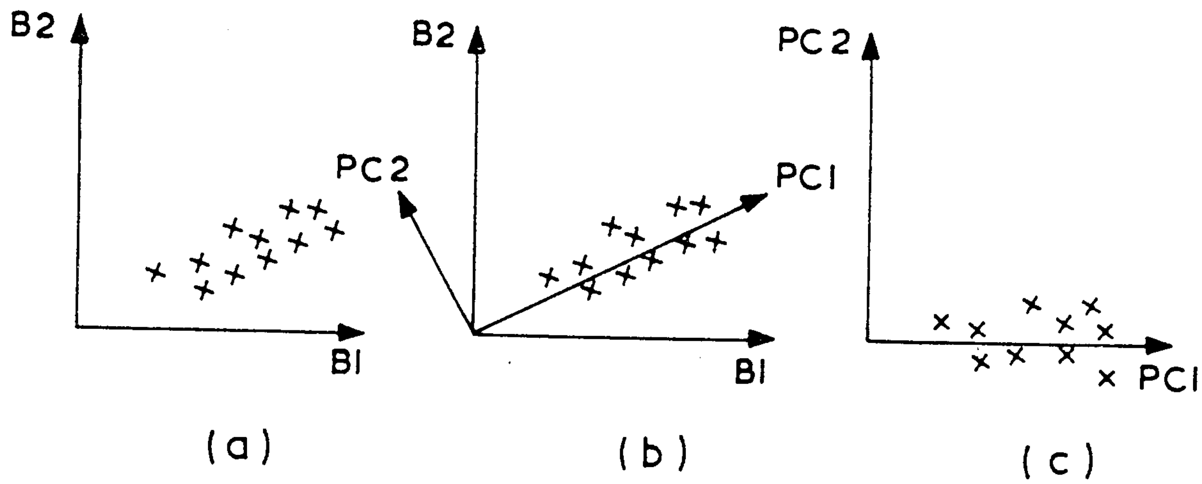


FIG: 4.10 PRINCIPAL COMPOMPONENT TRANSFORMATION IN 2 DIMENSIONS.

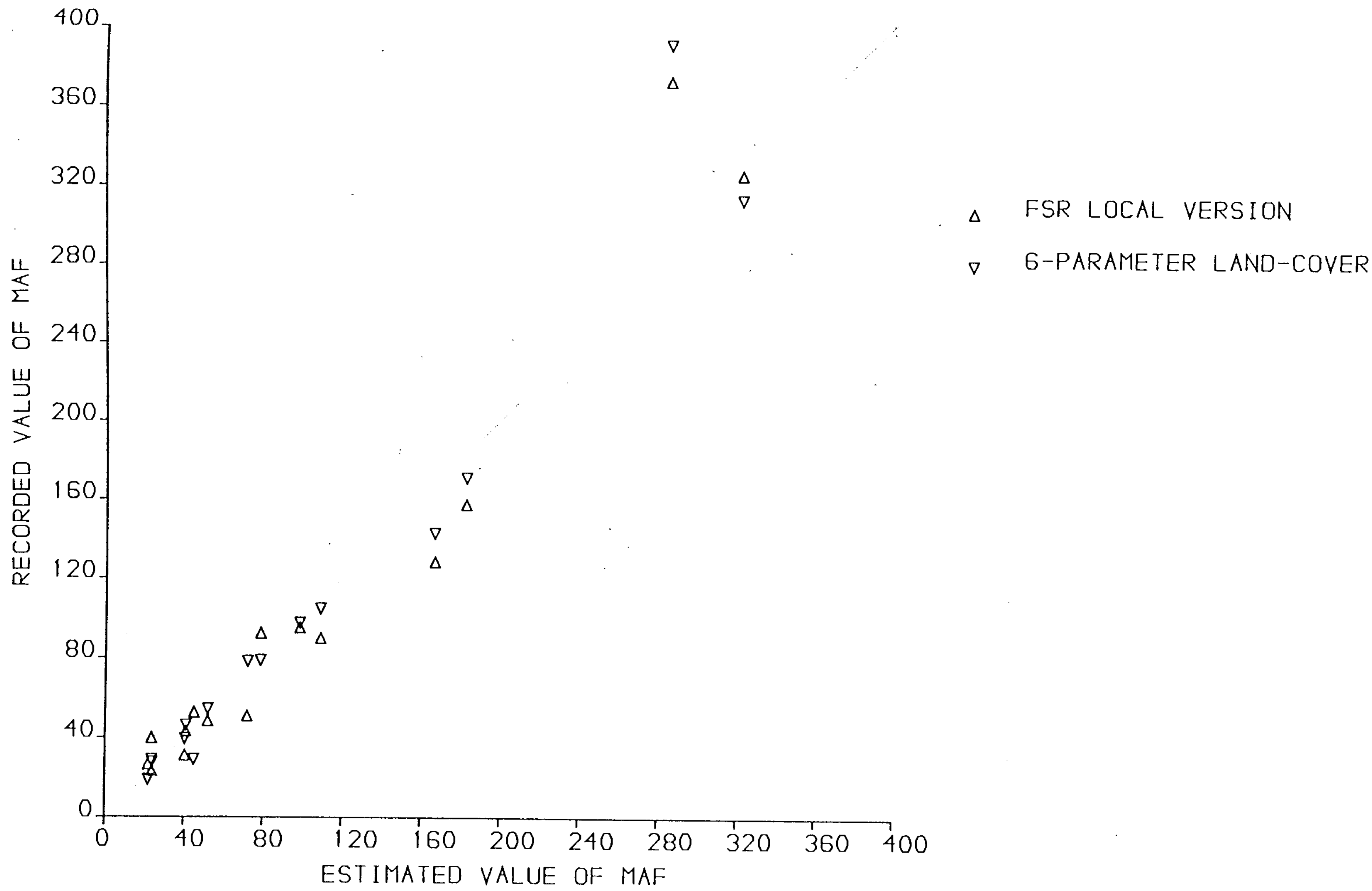


FIG. 6.1 COMPARISON OF ESTIMATES OF MEAN ANNUAL FLOOD

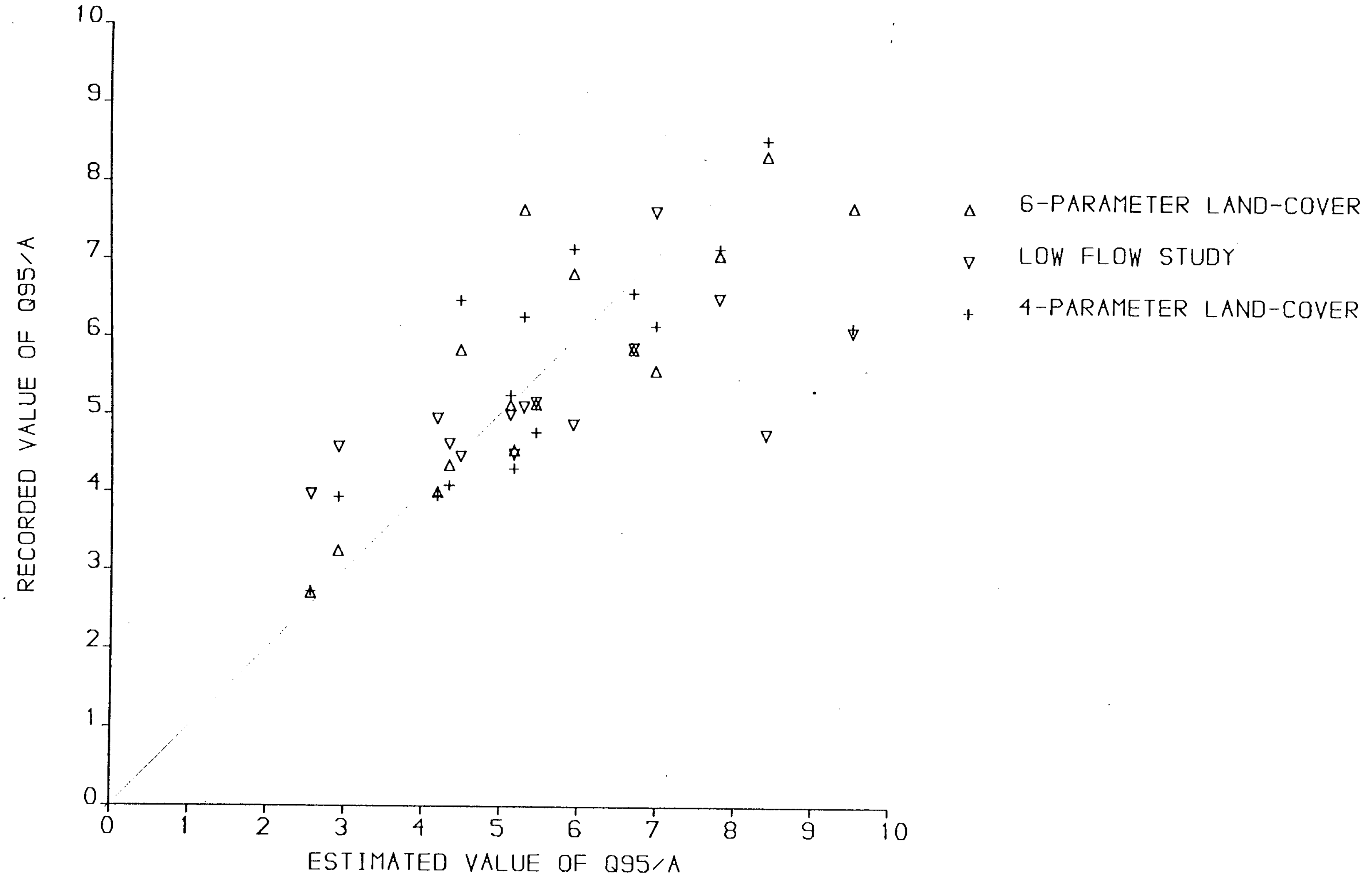


FIG. 6.2 COMPARISON OF ESTIMATES OF 95 PERCENTILE LOW FLOW

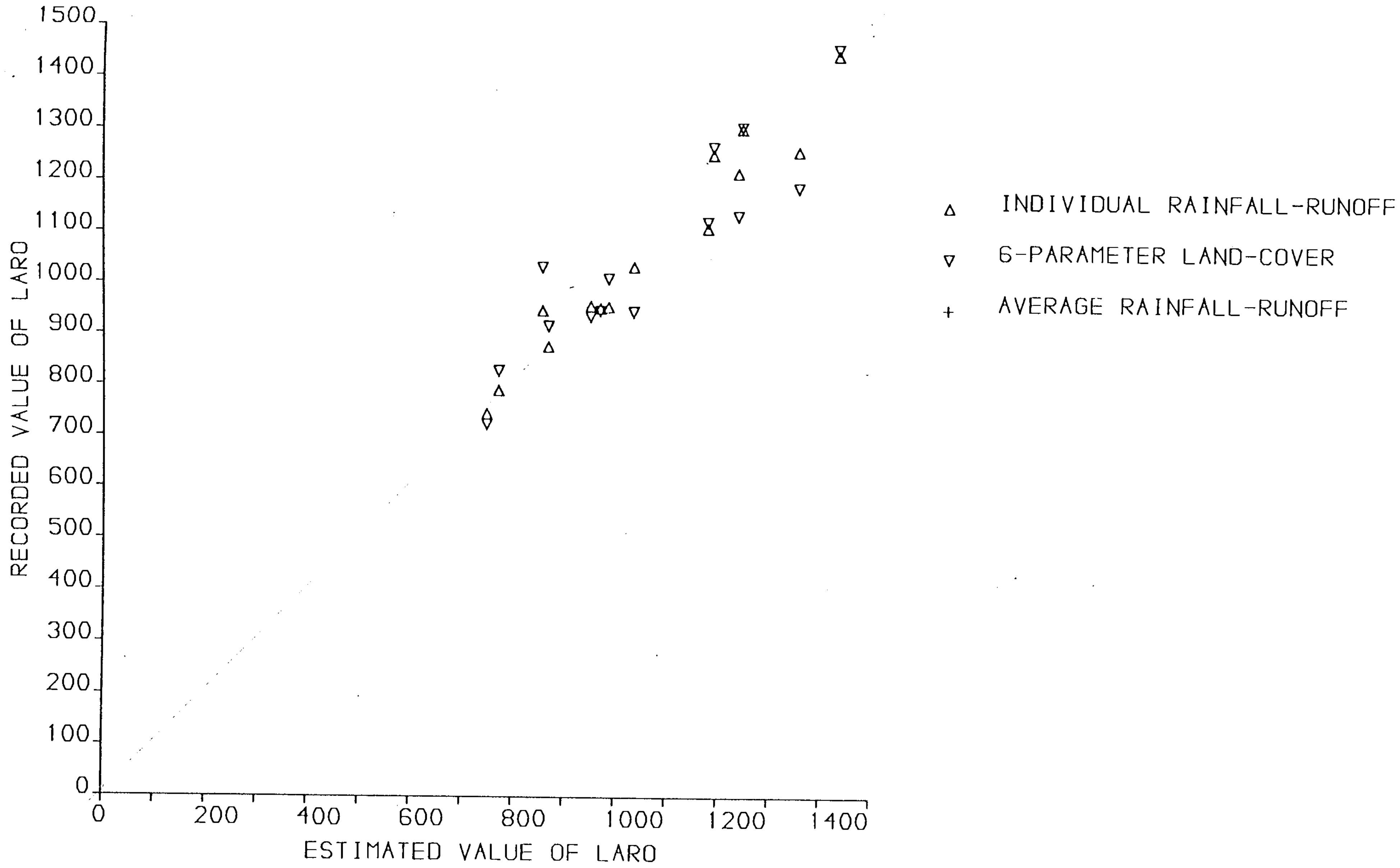


FIG. 6.3 COMPARISON OF ESTIMATES OF LONG AVERAGE RUNOFF

PART 4 SUMMARY AND CONCLUSIONS

PART 4 SUMMARY AND CONCLUSIONS

Underlying this study is the premise that remotely sensed data is holistic, bringing together many properties of the land surface, and so is likely to work well in models of surface processes. The particular aim of this study was to assess whether remotely sensed data could be interpreted in ways which would be useful in the assessment of the water resources of a region. This led to two quite distinct approaches, one based on the linear features within the scene and the other based on the thematic properties of the scene. In the first approach the data were interpreted to provide maps of the surface water features, such as rivers and lakes, and the geomorphic parameters that may be estimated from them, such as stream frequency. The second approach was to extract areas with similar properties, which were based either directly on the spectral data, or on more conventional properties which could be interpreted from the spectral data. The two approaches were brought together in regression models, in which key streamflow values were estimated on the basis of geomorphic and thematic parameters.

With regard to surface water mapping, it was concluded that visual interpretation was the only viable technique at present (1988), but that it was consistently reliable only in the case of major features such as main rivers. In using visual techniques it was found necessary to apply objective criteria for the identification of features, and to standardise procedures. Even when that was done results were found to be variable, in relation to quality of the image, topography, land-use and the subjectivity of the interpreter. Best results would be expected in well-dissected terrain, in uniform glassland, and in the absence of cultural features such as roads and railways.

Similarly, the automated computer based techniques of line-following which are currently being developed might be expected to work well in these ideal conditions. In more difficult conditions the interpretation depends on the level of intelligence and general knowledge which a human interpreter can bring to bear on the subject. Developments in artificial intelligence may go some way to overcoming that obstacle, but it seems unlikely that drainage mapping would ever offer sufficient rewards to make the proposition economically attractive.

Difficulties were found in judging the quality of drainage maps, and so an objective method of assessment was derived, which gives greater weighting to high order streams, and which is tolerant of small deviations in line.

With regard to thematic interpretations, two sets of catchment properties were derived, the first of which was based on the reflectance values in each band of the Landsat MSS data. Bands were examined singly, in ratios, and in terms of their principal components. They were found to be highly correlated, and it proved difficult to select any one spectral feature on statistical grounds alone. A small number of features were eventually selected on the basis of their physical relationships with either water or vegetation.

It was recognized that the statistical properties of spectral data depend on the illumination of the scene and on the sensor calibration. Thus, the statistics would be basically different in scenes of the same area on different dates, or scenes of adjacent areas on similar dates. It is possible to homogenise the data to some extent (as is done in the construction of mosaics of images) but to do so would be to modify the very data on which models are expected to operate. One way around this problem would be to use the primary data (spectral DN values) to interpret the traditional features of catchments(soils, land-use etc.) which would provide the necessary consistency.

A second set of properties was derived, which comprised a set of land-cover types, which could be identified from the spectral response of the land surface. The choice was largely determined by the experience of earlier researchers, but was modified in the light of a distinction between types of grassland which emerged during the analysis, and was found to be hydrologically significant for this study area. The classification of land-cover types using the clustering technique on several different images in turn was found to give inconsistent results, which highlighted the great difficulty of obtaining reliable, detailed ground-truth information for images which are not from recent times.

The streamflow record of a number of rivers in the study area was examined for features which would be useful to model. The conclusion was that the simple approach of estimating summary statistics was best, with the exception of catchment yield, which was best estimated using a rainfall-runoff relationship. Summary statistics covering the range from flood flows to low flows were extracted, and also the parameters of the annual rainfall-runoff relationships.

The correlation between streamflow statistics, spectral properties and land-cover types was examined, which gave a logical basis for the selection of parameters to use in regression models. The streamflow statistics were then regressed on the set of thematic properties and the best models were selected. The models were then extended to incorporate geomorphic parameters which could be estimated from visual interpretations.

It was recognised that much information on hydrologically significant features of the landscape reside in maps, which could be digitised and entered into Geo-Information Systems to enhance the regression models. However, the purpose of this study was to examine the benefit of satellite data, and so the models were not extended beyond that range.

It was found that streamflow was very strongly related to catchment area, so that simple but useful empirical equations for streamflow could be derived, for which accurate estimates of area could be made from visual interpretations. When the streamflow per unit area was examined it was found that there was little correlation with the spectral features of the scene. The conclusion was made that spectral characteristics would be significant only in areas of relatively natural vegetation. In developed areas the land-cover parameters would be preferable in regression models since the land-use would be important in determining the hydrology of the area.

Minimum daily flow was not modelled well, probably due to problems in gauging very low flows. However, the 95 percentile low flow, the mean annual flood, the long-average runoff and the parameters of the annual rainfall-runoff equation were all modelled well, and gave standard errors of estimate that were as good as or better than conventional models for the region.

It was recognized that there was a danger in developing models that fitted only the set of data used for their development. The set of catchment characteristics was reduced to a smaller, more generally applicable set, and the standard errors of estimate were seen to increase, but to remain within acceptable limits.

The principal conclusions are therefore :

(i) that the geology, soils, vegetation and land-use of an area combine to give a characteristic reflectance of solar energy, which can be measured remotely, and which a skilled interpreter can use in constructing maps of relevance to the water resources of an area.

(ii) that remotely sensed data can be employed in statistical models to predict the streamflow characteristics of ungauged catchments, with an acceptable accuracy,

(iii) that such models may easily be updated to reflect changing conditions, and

(iv) that such models may be made sufficiently general and robust that they may be recalibrated for use in regions other than the one in which they were originally developed.

In carrying out this work the body of research has been extended in the following areas :

(i) in the visual interpretation of surface water features the problems of perception and subjectivity have been addressed, and an objective method of assessment has been devised. The variety of geographic areas used for trials has been extended.

(ii) the importance of purely spectral information has been recognized, and attempts have been made to use such information in place of conventional map data.

(iii) the scope and range of regression models has been greatly increased, and they have been tested against a greater range of conventional models.

PART 4 APPLICATIONS

The practical relevance of the work described above lies in the evaluation of the surface water resources of a region. In Parts 2 and 3 the two key issues of mapping and modelling were examined separately. In this section those ideas will be brought together in the context of carrying out an appraisal of the surface water resources of a previously un-surveyed area.

The first step in the survey would be to map the rivers and lakes, and it has been shown that a visual interpretation of the infrared imagery is capable of providing reliable maps of principal drainage lines, sufficient to demarcate the main drainage basins, and to provide a guide to catchment size so that the planning of a hydrometric network could be carried out. (It seems likely that in undeveloped areas the reliability of interpretation at this level will be considerably better than that achieved for the South Wales study area, due to the absence of cultural features which may be confused with rivers). The techniques of visual interpretation are appropriate for many areas of the developing world where skills in air-photo interpretation already exist and could be extended to satellite image interpretation. An improvement in detail could be achieved by the use of existing aerial photography or other documents, but only at the expense of some loss of consistency in mapping.

Following the initial reconnaissance and mapping, the survey would normally proceed with the establishment of a small number of meteorological and river gauging stations, sufficient to gain some idea of the rainfall and streamflow in a few key catchments. Within a hydrologically homogeneous region it would then be possible to calculate simple models of the type :

$$Q = c. \text{AREA}$$

and $$\text{Yield} = a + b. \text{LARF}$$

Having once established the values of the parameters a, b and c from the gauged catchments the models could be used throughout the region, on the assumption that the hydrological response is similar throughout the region i.e. that geology, soils, vegetation and climate are uniform. Visual interpretation of satellite imagery could provide justification for this assumption, and the simple visual techniques of mapping could provide sufficiently reliable estimates of catchment area needed to derive the streamflow estimates.

Simple models of this kind can be very effective within a homogeneous region, but cannot be relied on to give good estimates for catchments outside the region of gauging, or for regions within which hydrologic conditions vary. When heterogeneity exists then the values of the parameters a, b and c may be different for every catchment.

Suitable regression models may be developed for unit streamflows and the parameters of the yield equation, based on the spectral characteristics of the catchments i.e.

$$Q/A = f \text{ (spectral characteristics)}$$

The technique now is limited by the accuracy of the drainage mapping, and such predictive equations are likely to be of value only to the largest catchments. The location and delineation of smaller catchments using visual interpretation is unlikely to be adequate.

When mapping at say 1:250,000 scale becomes reliable, then the possibility exists for registering the imagery to map coordinates, and locating catchment boundaries on the image to an acceptable accuracy. The models may then be refined, and as development takes place fresh imagery will provide the means for updating hydrological models in step with changing land-use. In this way the techniques remain of practical value, keeping pace with development, until more sophisticated models are required, at which point the techniques may still be of value in estimating values of the parameters of models of individual processes.

PART 4 RECOMMENDATIONS FOR FUTURE WORK

The argument made in preceding sections proposes that models based solely on spectral parameters should work well in areas of natural vegetation. It was not possible to test this idea in the region for which data was available in this project, and so it is recommended that a similar study be undertaken for a region of unmodified vegetation. Unfortunately, it is usually the case that such areas are in the developing world where data are required rather than available.

An important extension of the investigation would be into catchments of small size. Primarily this would be useful to determine the minimum size of catchment that can reliably be modelled using the fairly coarse resolution data that satellite systems provide. Again, the problem is that very few catchments of small size are gauged, and so relevant data would be hard to acquire. An alternative would be to conduct a theoretical sensitivity analysis. The importance of this aspect is that if reliability can be demonstrated for small catchments, then the way is opened for the development of remote sensing inputs into partially distributed models.

Looking at the development of satellite hydrology in a more general way, it can be seen that benefits can be gained from two aspects of satellite data :

(i) that it is repetitive

(ii) that it is spatially distributed

Work has already been done to exploit the frequent coverage of meteorological satellites to make rainfall estimates for parts of the world where gauging networks are inadequate. Good quality estimates may now be made for periods down to 1 month, and have been used successfully in revising yield estimates (e.g. Barrett,1970). Shorter duration rainfall estimates are susceptible to greater error, but are useful in drought monitoring (e.g. Barrett, 1978) Coarse resolution, highly repetitive data from the AVHRR satellites have also been used to provide biomass estimates which give an indication of water demand, and the potential for runoff (Tucker et al 1984) Thus, the potential exists for the integration of such techniques to provide regular, up-dated estimates of the state of catchments (e.g. soil moisture deficit) that would be of value in real-time forecasting of streamflows for reservoir operation and flood forecasting.

The techniques examined here have been concerned with the modification of rainfall to produce run-off in various forms. The logical complement of these models is the use of satellite data to estimate the proportion of rainfall which is transmitted via infiltration and percolation to recharge acquifers. An approach which appears to be a fairly simple extension of the work done here is to make classifications of land-cover, which may be used to distinguish areas of recharge, and then to make classifications of vegetation type,which would provide the basis for the estimation of evapo-transpiration losses, and hence the magnitude of the recharge. A particular benefit of satellite data is that it is spatially organised, and capable of being fitted to the grid layout typical of groundwater models.

A major limitation of the use of satellite imagery is in the size of the scene (typically 185kms wide) which is small in relation to the normal scale of regional studies. Whilst it is possible to construct mosaics of images, the orbiting patterns of the satellites makes it very difficult to obtain good quality images of adjacent scenes which are reasonably close in time. If the best of the available images are used then it is likely that adjacent scenes will be of different seasons and even different years, in which case the hydrological state of the catchment and the distribution of vegetation types will be different. Thus data or interpretation taken from the two adjacent images will be heterogeneous to some degree, and it is recommended that investigations be made of the degree of heterogeneity which is likely to exist, and the development of techniques to overcome the problem.

The most important avenue for further work undoubtedly lies in the integration of many forms of data in spatially organized data bases

(Geo-Information Systems) which will provide the foundation for many varied environmental models. The release in 1988 of digital elevation data by the Ordnance Survey of Great Britain has made it possible to construct data bases on a 50m grid, referenced to the National Grid, which can store layers of data such as elevation, soil type, land-cover (from image interpretation), drainage networks, communications etc. It appears likely that progress in the handling of spatial information is going to determine to a large extent the type of model which will be of most practical value in the future, and that the choice of model will no longer determine the data requirements as has normally been the case in the past.

The value of spatially organised, multi-layered data is too great to ignore, and it is recommended that the next stage of the work in this field should be the development of models which exploit these data bases. Techniques will have to be developed for the integration of vector information (drainage nets, catchment boundaries) and raster information (i.e. all thematic data), and for the modelling of processes using parameters which may be estimated from the data which are available. The implication is that such models will be spatially complex, but conceptually simple.

REFERENCES

ACKERMAN, E. (1936) Dambos in Nordrhodesia. Wiss. Veroff, Leipzig, 4, 149-157.

ASTARAS, T. (1985), Drainage network analysis of Landsat images of the Olympus-Pieria mountain area of Northern Greece. Int.J. Remote Sensing, 6-5, 673-686.

BAKER, V.R. (1982). The Channels of Mars, Hilger.

BARRETT, E.C. (1970). The estimation of monthly rainfall from satellite data. Mon. Weather Rev. 98, 322-327.

BARRETT, E.C. (1978). The use of weather satellite data in the evaluation of national water resources. Space Research, 19, 41-46.

BELLAVIA, G. amd ELGY, J. (1986). Spatial feature extraction from radar imagery. Int. Archives of Photogrammetry and Remote Sensing, 26, 7/1, 99-102.

BISWAS, A.K. (1971), History of Hydrology, North-Holland, Amsterdam.

BLANCHARD, B. (1975). Measuring watershed runoff capability with ERTS data. In. Proc 3rd ERTS-1 Symp. ed. NASA, USA, 1089-1098.

BOX, G.E.P. and JENKINS, G.M. (1970). Time Series Analysis : Forecasting and Control, Holden-Day, San Francisco.

CAMPBELL, J.B. (1987). Introduction to Remote Sensing, Guildford Press. UK.

CLARK, M.J., GURNELL, A.M., MILTON, E.J. (1986), Remote Sensing of heath vegetation and its significance in hydrological modelling. Inst. Brit. Geog. Ann. Conf., Reading, 1986.

CRAWFORD, N.H. and LINSLEY, R.K. (1966). Digital Simulation in Hydrology: Stanford Watershed Model IV, Stanford University, Dept. Civ. Eng. Tech. Report 39, Stanford.

DEUTSCH, M. and RUGGLES, F.M. (1977). Hydrological applications of Landsat imagery of the 1973 Indus River flood. Proc. 11th Int. Symp. on Remote Sensing of the Environment, ERIM, 1, 1373-1375.

DRAYTON, R.S., KIDD, C.H.R., MANDEVILLE, A.N. and MILLER, J.B. (1980). A Regional Analysis of River Floods and Lowflows in Malawi, Report No. 72 Institute of Hydrology, Wallingford, Oxon.

FAO, (1982). Remote Sensing applied to renewable resources. RSC series, 6.

FAULKNER, R.D. (1986). Personal communication.

FIERING, M. (1967). Streamflow Synthesis, Harvard University Press, Cambridge, Mass., USA.

FLEMING, G. (1975). Computer simulation techniques, Elsevier, New York.

FOWLER, T.R., CASTRUCCIO, P.A. and LOATS, H.L. (1977). The application of remote sensing to the development and formulation of hydrologic models. In : NASA final report No. CR - 150326.

GREGORY, K.J. (1986). Personal communication.

GROVES, J.R. and RAGAN, R.M. (1983). Development of a remote sensing based continuous streamflow model. In: Proc. 17th Int. Symp. on Remote Sensing of the Environment, Ann Arbor, Michigan, 447-456.

GUMBEL, E.J. (1941). The return period of flood flows. Ann Math Stat, 12, 2-163.

GUMBEL, E.J. (1958). Statistics of Extremes, Columbia University Press, New York.

GURNEY, C.M. and TOWNSHEND, J.R.G. (1983). The use of contextual information in the classification of remotely sensed data. Photogram. Eng. and Rem. Sens., 49-1, 55-64.

HARDY, D.D. (1985). Today, the earth is 100 000 000 000 000 bits. Int.Conf. on Advanced Technology for Monitoring and Processing Global Environmental Data, Remote Sensing Society, Reading, 1-4.

HAZEN, A. (1914). Storage to be provided in impounding reservoirs for municipal water supply. ASCE, 77, 1539-1669.

HAZEN, A. (1930). Flood Flows, John Wiley and Sons, New York.

HOLLYDAY, E.F. and PLUHOWSKY, E.J. (1976). Improving estimates of streamflow characteristics. In: ERTS-1: A new window on our planet, USGS, Water Supply Paper 929, 136-137.

HORTON, R.E. (1933). The Role of Infiltration in the Hydrologic Cycle. Trans.Am.Geophys. Union, 13, 446-460.

HORTON, R.E. (1932). Drainage basin characteristics. Trans.Am.Geophys, Union, 14, 350-361.

HUFSCHMIDT, M. and FIERING, M. (1962). Simulation Techniques for the Design of Water Resources Systems, Harvard University Press, Cambridge Mass.

HURST, H.E. (1951). Long Term Storage capacity of Reservoirs. Trans. ASCE, 116, 770-808.

HUTTON, S.M. and DINCER, T. (1979). Using Landsat imagery to study the Okavango swamp, Botswana. Proc. 5th Ann.Symp. on Remote Sensing of the Environment, AWRA.

INSTITUTE OF HYDROLOGY, (1980). Low Flow Report, Institute of Hydrology, Wallingford, UK.

JACKSON, T.J. and RAWLS, W.J. (1981). SCS Curve Numbers from a Landsat data base. Water Res. Bull., 17(5), 857-862.

JACKSON, T.J., RAGAN, R.M. and FITCH, W.N. (1977). Test of Landsat based urban hydrological monitoring. J.W.R.M. Div. ASCE, 103(WR 1). 141-158.

JACKSON, T.J., RAGAN, R.M. and SHUBINSKI, R.P. (1976). Flood frequency studies on ungauged watersheds using remotely sensed data. Proc. Natl. Symp. on Urban Hydrology, Lexington, Kentucky, 31-39.

KILLPACK, D.P. and McCOY, R.M. (1981). An application of Landsat derived data to a regional hydrological model. Remote Sens. Quart., 3(2), 27-33.

LANE, W. (1979). Applied Stochastic Techniques, User's Manual, U.S. Bureau of Reclamation, Denver, Colorado.

LAWRANCE, A.J. amd KOTTEGODA, N.T. (1977). Stochastic Modelling of Riverflow Time Series. J.Roy.Stat.Soc., A Pt. 1, 1-31.

LETTENMAIER, D.P. and BURGES, S. J. (1977). Operational Assessment of Hydrologic Models of Long Term Persistence. Water Res. Res., 13,1, 113-124.

LIGHTHILL, M.J. and WHITHAM, G.B. (1955). On Kinematic Waves: 1-Flood Movement in Long Rivers. Proc.Roy.Soc., 229 Series A, 281.

LINSLEY, R.K. and CRAWFORD, N.H. (1960). Computation of a Synthetic Streamflow Record on a Digital Computer, Int.Assoc.Sci.Hydrol. Publ. 51, 526-538.

LINSLEY, R.K., KOHLER, M.A. and PAULHUS, J.L.H. (1949). Applied Hydrology, McGraw-Hill, New York.

LINSLEY, R.K., KOHLER, M.A. and PAULHUS, J.L.H. (1982). Hydrology for Engineers, McGraw-Hill, Kogakusha.

MACKEL, R. (1985). Dambos and related landforms in Africa. J. Geomorphol, (Berlin), 52, 1-23.

MANDELBROT, B.B. and WALLIS, J.P. (1969) Computer Experiments with Fractional Gaussian Noise. Water Res. Res., 8, 228-267.

MATALAS, N.C., (1967). Mathematical Assessment of Synthetic Hydrology. Water Res. Res., 3,4, 937-945.

MEJIA, M.I., RODRIGUEZ, I.I. and DAWDY, D.R. (1972). Streamflow Simulation: The Broken Line Process as a Potential Model for Hydrologic Simulation. Water Res. Res., 8, 931-941.

MOORE, R.V. (1983). Digitising the UK River network. In: Auto-Carto 6: Proc. 6th Int.Symp. on Automated Cartography, Ottawa, 2.288-2.298.

MOORE, R.V., MORRIS,D.G.. and BONVOISIN, N.J. (1987). Digital mapping as an aid to hydrological analysis. Proc. 1st Ann.Symp. British Hydrological Society, Hull, 31.1-31.13.

MORRIS, D.G. and HEERDEGEN, R.G. (1987). Catchment and channel network definition from gridded ground elevation data. (In press).

NASA, (1977). Detection and mapping (dam) package. NASA news release, 76-4.

NASH, J.E. (1960).A unit hydrograph study with particular reference to British catchments. Proc. Inst.Civil Engrs., 17, 249-282.

NERC, (1975). Flood Studies Report, Natural Environment Research Council, Swindon, UK.

O'CONNELL, P.E. (1974). A simple stochastic modelling of Hurst's Law. Publn. No. 100, IAHS, 169-187.

PEACEGOOD, G. and WILKINSON, G.G. (1985). A knowledge-based system for the identification of linear features in satellite images. Proc. Workshop on Image Understanding, University College, London.

PENMAN, H.L. (1948). Natural Evaporation from Open Water, Bare Soils and Grass. Proc. Royal Soc., 193, Series A, 120-145.

PENMAN, H.L. (1956). Estimating Evaporation. Trans. Am. Geophys. Union, 37, 43-50.

RAGAN, R.M. and JACKSON, T.J. (1975). Use of satellite data in urban hydrological models. J.Hyd.Div., ASCE, 101 (HY 12), 1469-1475.

RAGAN, R.M. and JACKSON, T.J. (1980). Runoff synthesis using Landsat and SCS models. J.Hyd.Div., ASCE, 106 (HY 5), 667-678.

RANGO, A., FELDMAN, A., GEORGE, T.S. and RAGAN, R.M. (1983). Effective use of Landsat data in hydrological models. Water Res. Bull. 19(2), 165-174.

RANGO, A., FOSTER, J. and SALOMONSON, V.V.V., (1975). Extraction and utilization of space acquired physiographic data for water resources development. Water Res. Bull., 11(6), 1245-1255.

SABINS, F.F. (1987). Remote Sensing Principles and Interpretations, Freeman, New York.

SALOMONSON, V.V., AMBARUCH, R., RANGO, A. and ORMSBY, J.P. (1975). Remote sensing requirements as suggested by watershed model sensitivity analyses. 10th. Int. Symp. on Remote Sensing of the Environment, ERIM, Ann Arbor, Michigan. 1273-1284.

SCHNEIDER, S.R., McGINNIS, D.F. and PRITCHARD, J.A. (1979). Use of satellite infrared data for geomorphology studies. Rem. Sens. of Environment, 8, 313-330.

SEIDEL, K., ADE, F. and LICHTENEGGER, J. (1983). Augmenting Landsat MSS data with topographic information. IEEE Trans. Geosc. and Rem. Sens., GE-21, 3, 252-258.

SHREVE, R.L. (1966). Statistical law of stream numbers. J. Geol., 74, 17-37.

SLACK, R.B. and WELCH, R. (1980). SCS runoff curve number estimates from Landsat data. Water Res. Bull., 16(5), 887-893.

SNYDER, F.F. (1938). Synthetic Unit Graphs. Trans. Am.Geophys. Un., 19, 447.

THEIS, C.V. (1935). The Relation between the Lowering of the Piezometric Surface and the Rate and Duration of Discharge of a Well using Groundwater Storage. Trans.Am.Geophys. Union, 16, 519-524.

THOMAS, H.A. and FIERING, M.B. (1962). Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. Chapter 12, Design of Water Resources Systems, ed. A. Maass et al, Harvard University Press, Cambridge, Mass.

THORNTHWAITE, C.W. (1944). Report of Committee on Transpiration and Evaporation. Trans.Am.Geophys. Union, 25, part 5, 687.

TOMLINSON, R.F. (1984). Geographic Information Systems - a new frontier. Proc.Int.Symp. on Spatial Data Handling, Zurich.

TOWNSHEND, J.R.G. (1981). Terrain Analysis and Remote Sensing, Allen and Unwin.

TUCKER, C.J., GATLIN, J.A. and SCHNEIDER, S.R. (1984). Monitoring vegetation of the Nile delta with NOAA-6 and NOAA-7 AVHRR imagery. Photog.Eng. and Remote Sensing, 50, 53-61.

US SOIL CONSERVATION SERVICE, (1972). National Engineering Handbook, Section 4, Hydrology, US Govt. Printing Office, Washington, DC.

USDA SOIL CONSERVATION SERVICE, (1975). Urban Hydrology for Small Watersheds, USDA, Tech. Release No. 55, USA.

VOSS, A.W., BAKER, J.E., HAUSER, G.E. and NEWTON, D.W. (1978). The use of Landsat-derived land cover data in a flood peak correlation study. Proc. 44th Meeting American Soc. Photogrammetry, 135-146.

WANG, S., ELLIOTT, D.B., CAMPBELL, J.B., ERICH, R.W. and HARALICK, R.M. (1983). Spational reasoning in remotely sensed data. IEEE Trans. Geosc. and Rem. Sens., GE-21, 1, 94-101.

WEBB, R.P., CARMAK, R. and FELDMAN, A., (1980). Determination of land use from satellite imagery for input into hydrologic models. Proc. 14th Int. Symp. on Remote Sensing of the Environment, Environ. Res. Inst. Michigan, USA.

WORLD METEOROLOGICAL ORGANIZATION, (1974). Guide to Hydrological Practices, WMO Publ. No. 168, Geneva.

YOUNG, J.A.T., (1986). A UK Geographic Information System. Monograph No. 1, Remote Sensing Society, Nottingham, UK.

Appendix 1 - Hydrographs

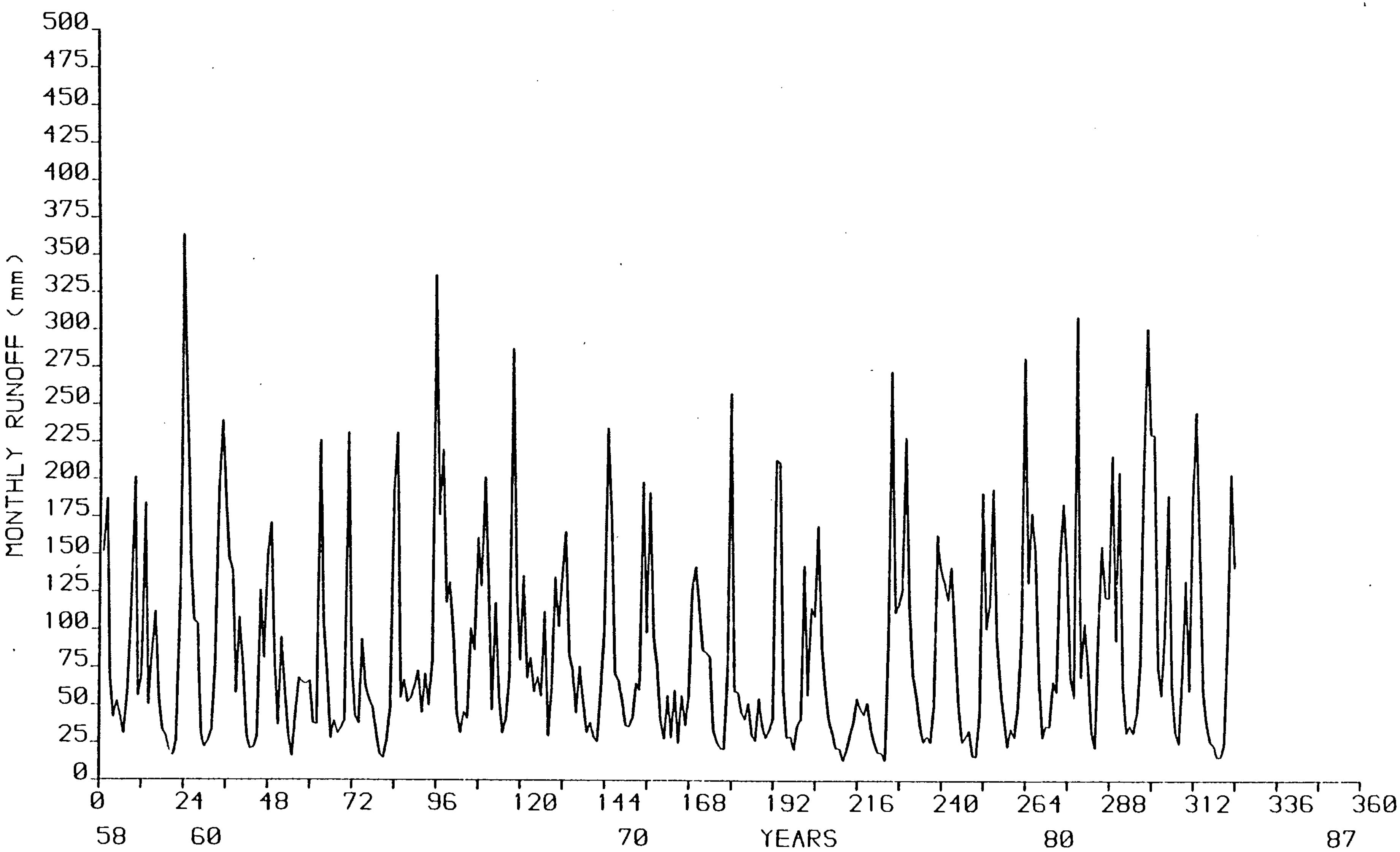


FIG. MONTHLY RUNOFF, RIVER EBBW

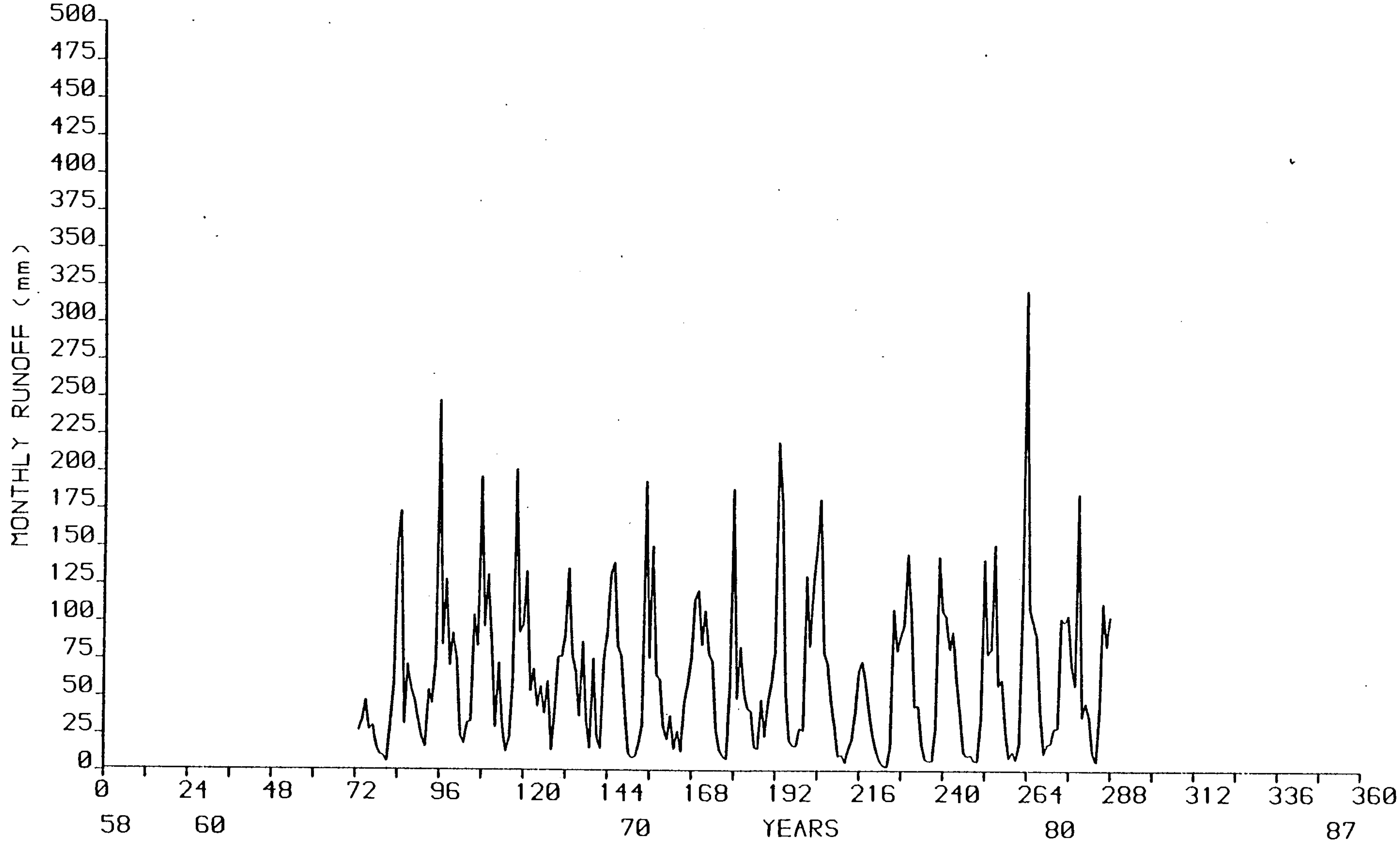


FIG. MONTHLY RUNOFF, RIVER HONDDU

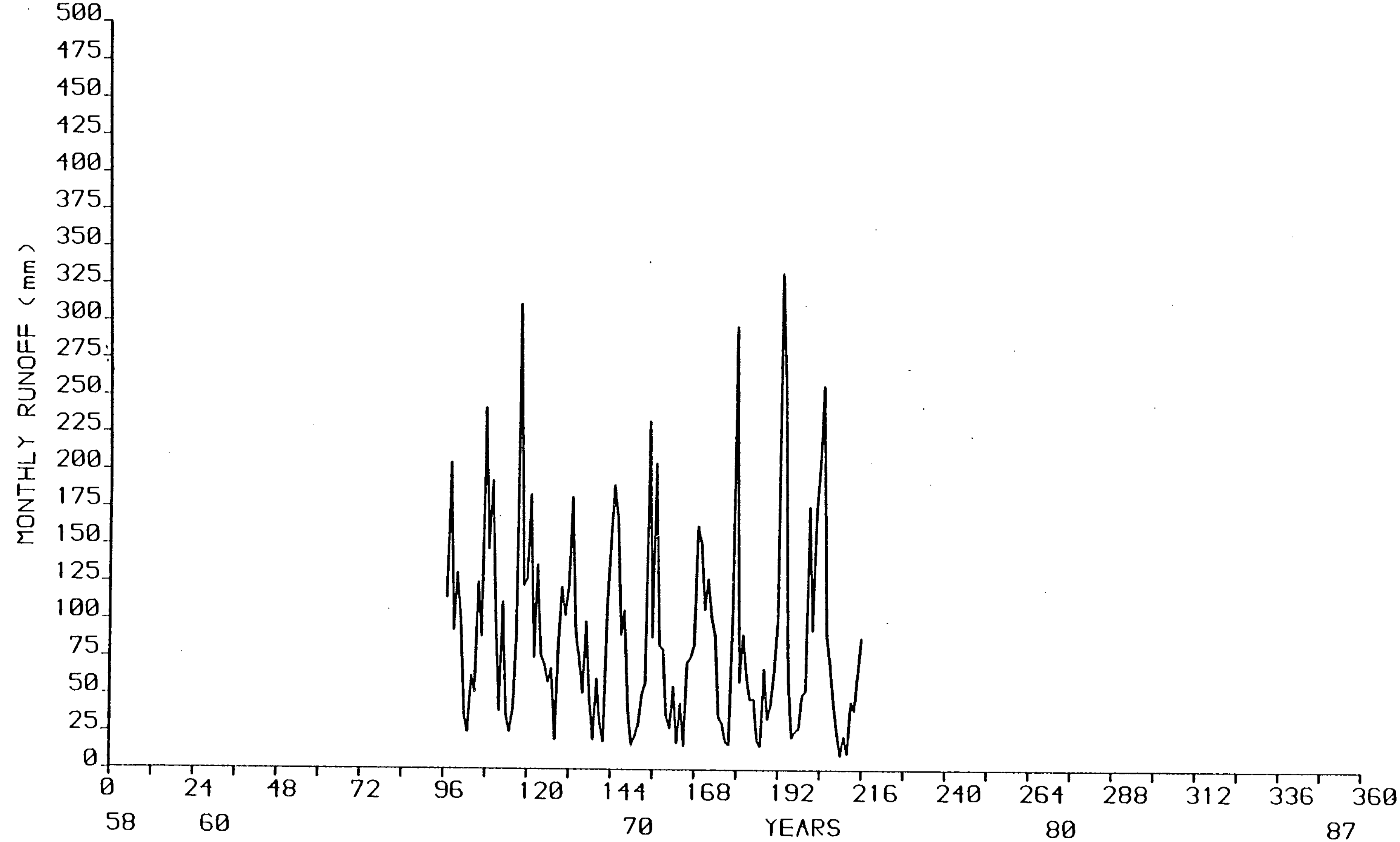


FIG. MONTHLY RUNOFF, RIVER USK AT LLANDETTY (NATURALISED)

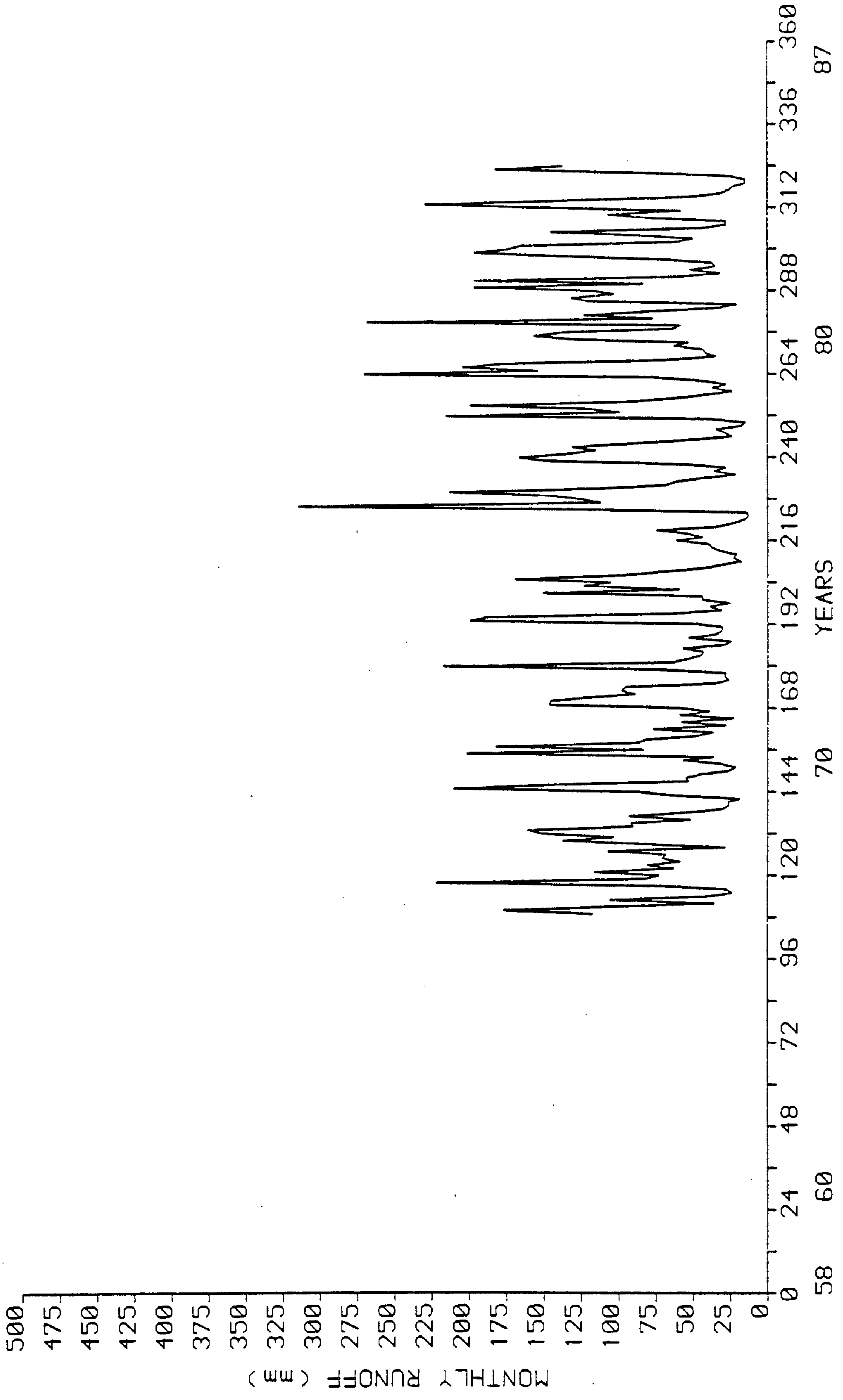


FIG. MONTHLY RUNOFF, AFON LWYD

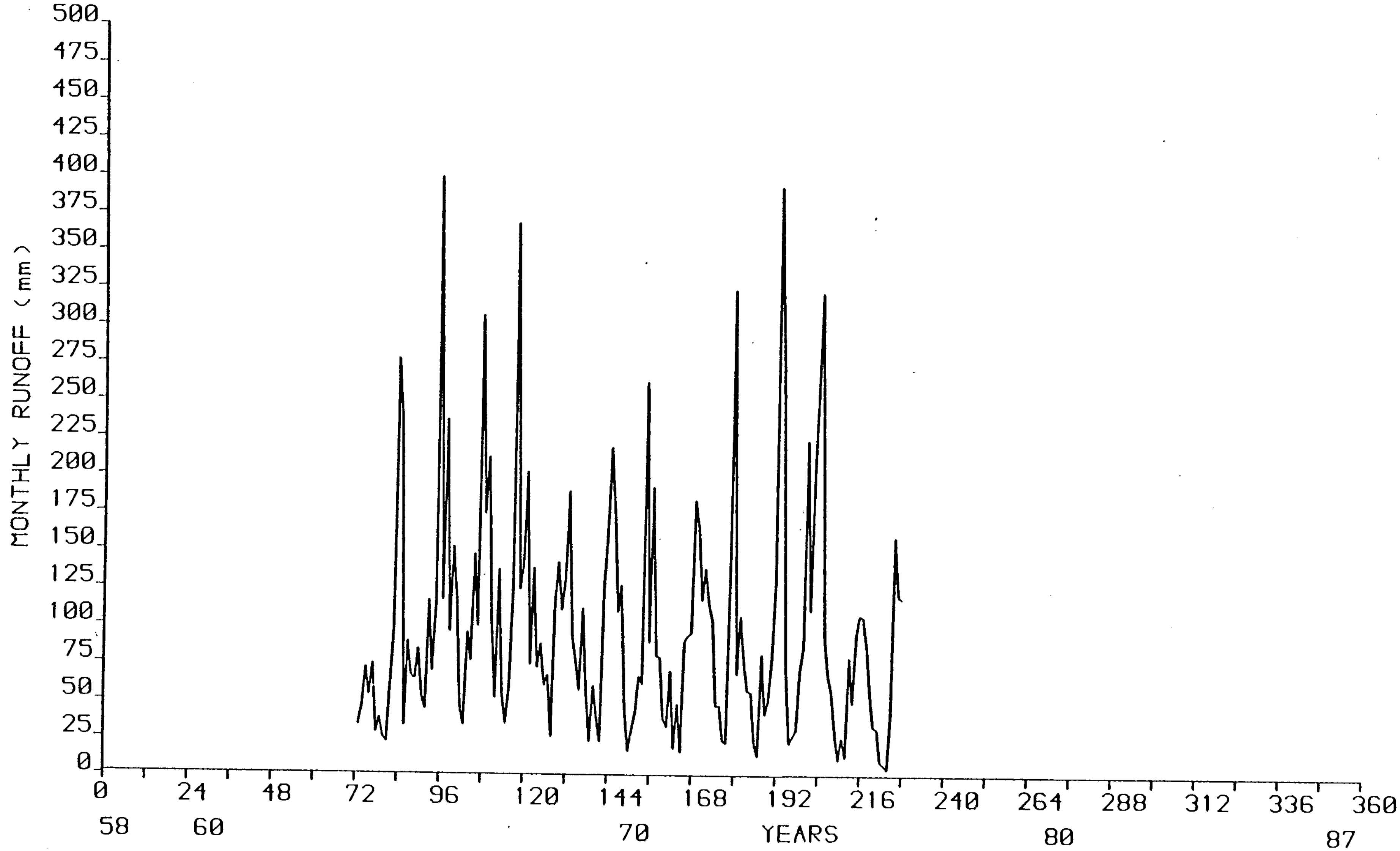


FIG. MONTHLY RUNOFF, RIVER USK AT TRALLONG (NATURALISED)

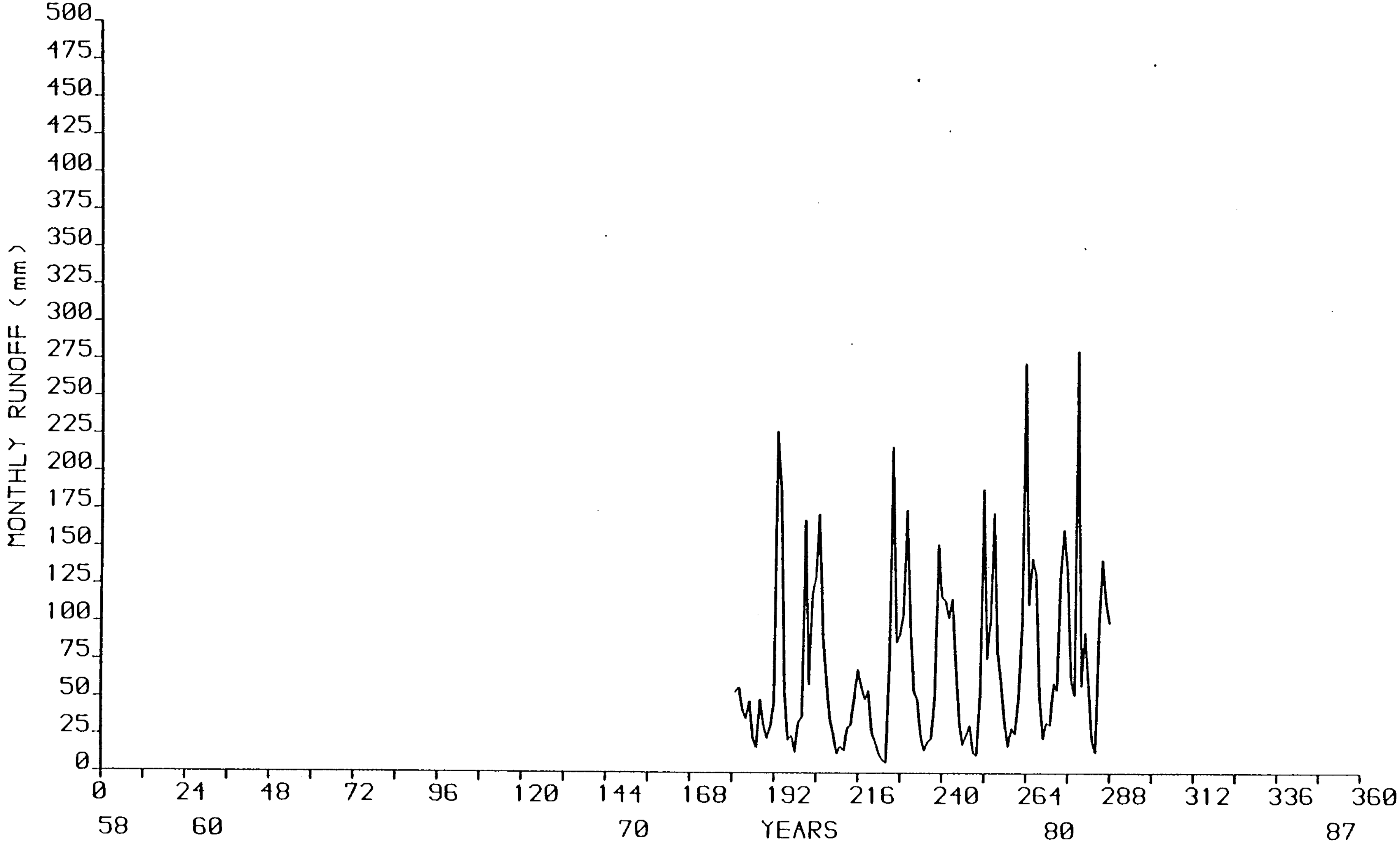


FIG. MONTHLY RUNOFF, RIVER SIRHOWY

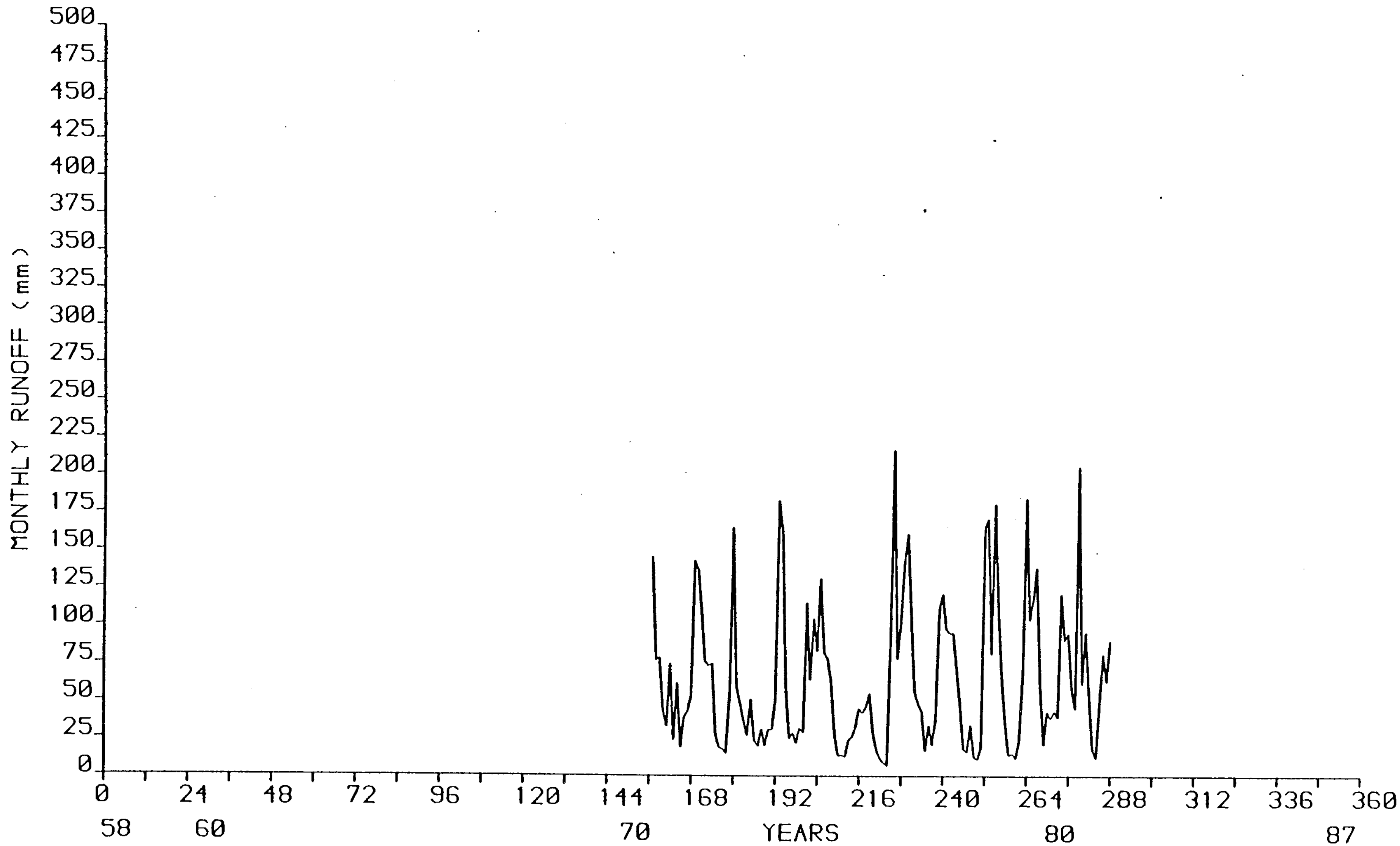


FIG. MONTHLY RUNOFF, RIVER GRWYNE

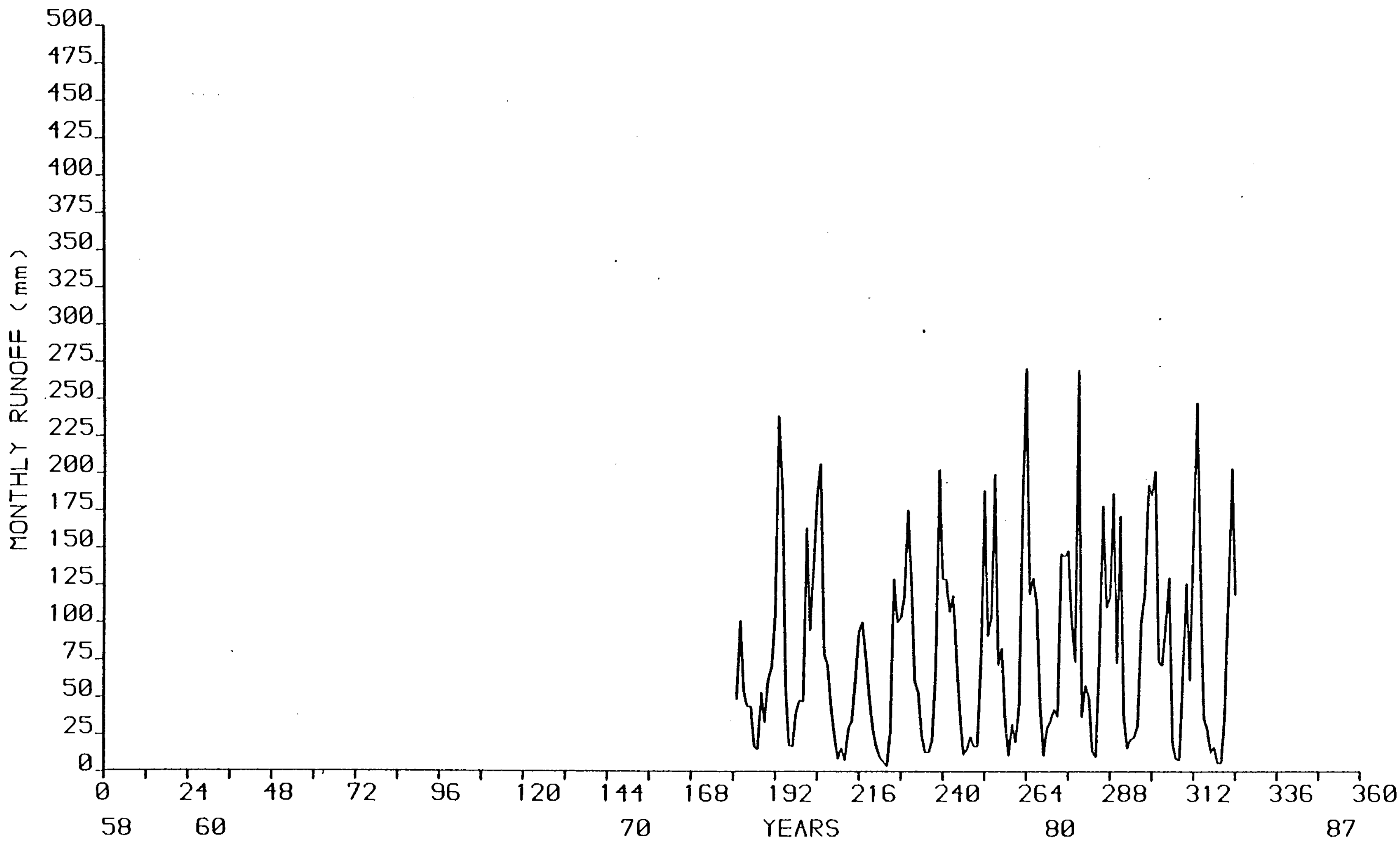


FIG. MONTHLY RUNOFF, RIVER YSCIR

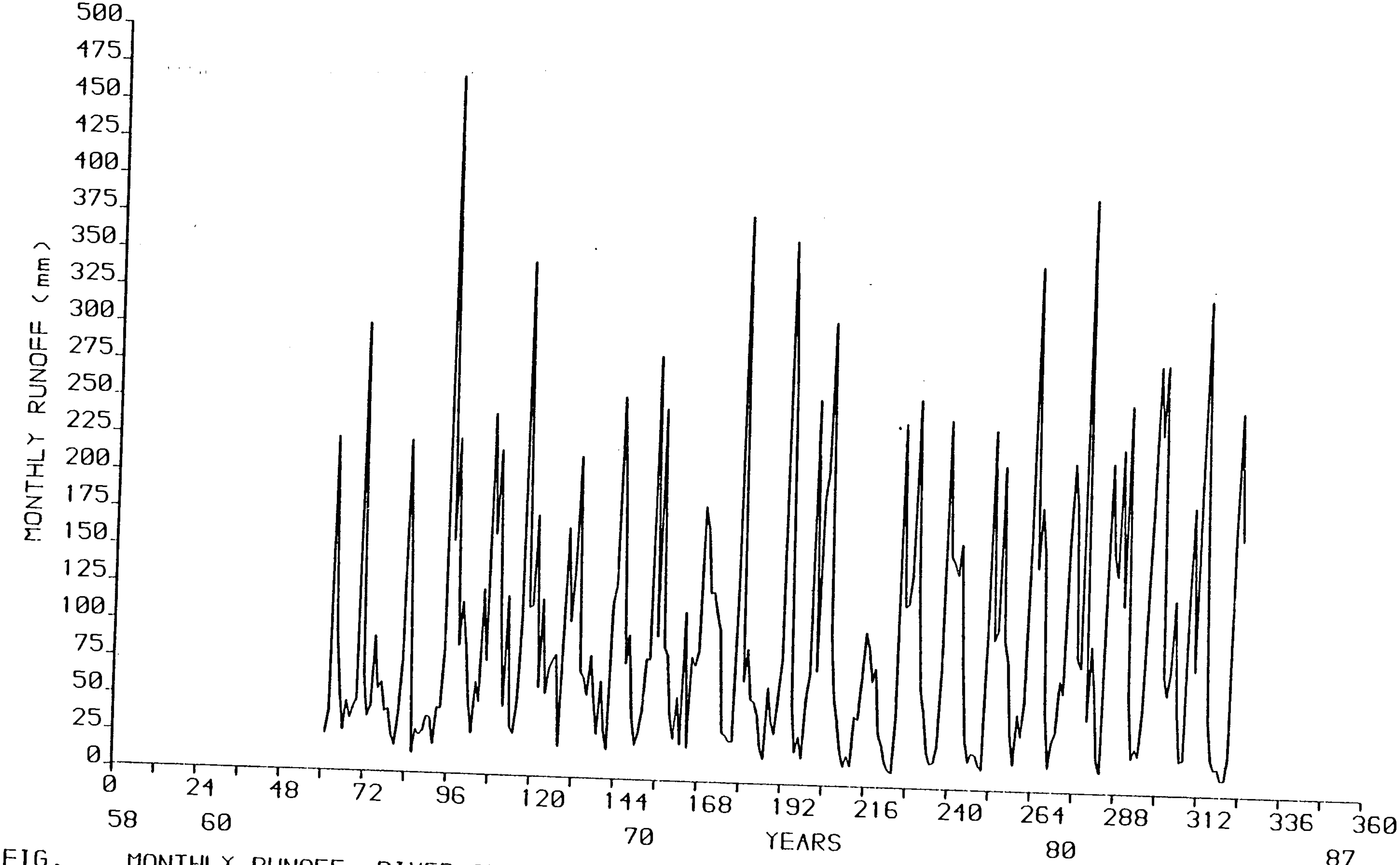


FIG. MONTHLY RUNOFF, RIVER CYNON

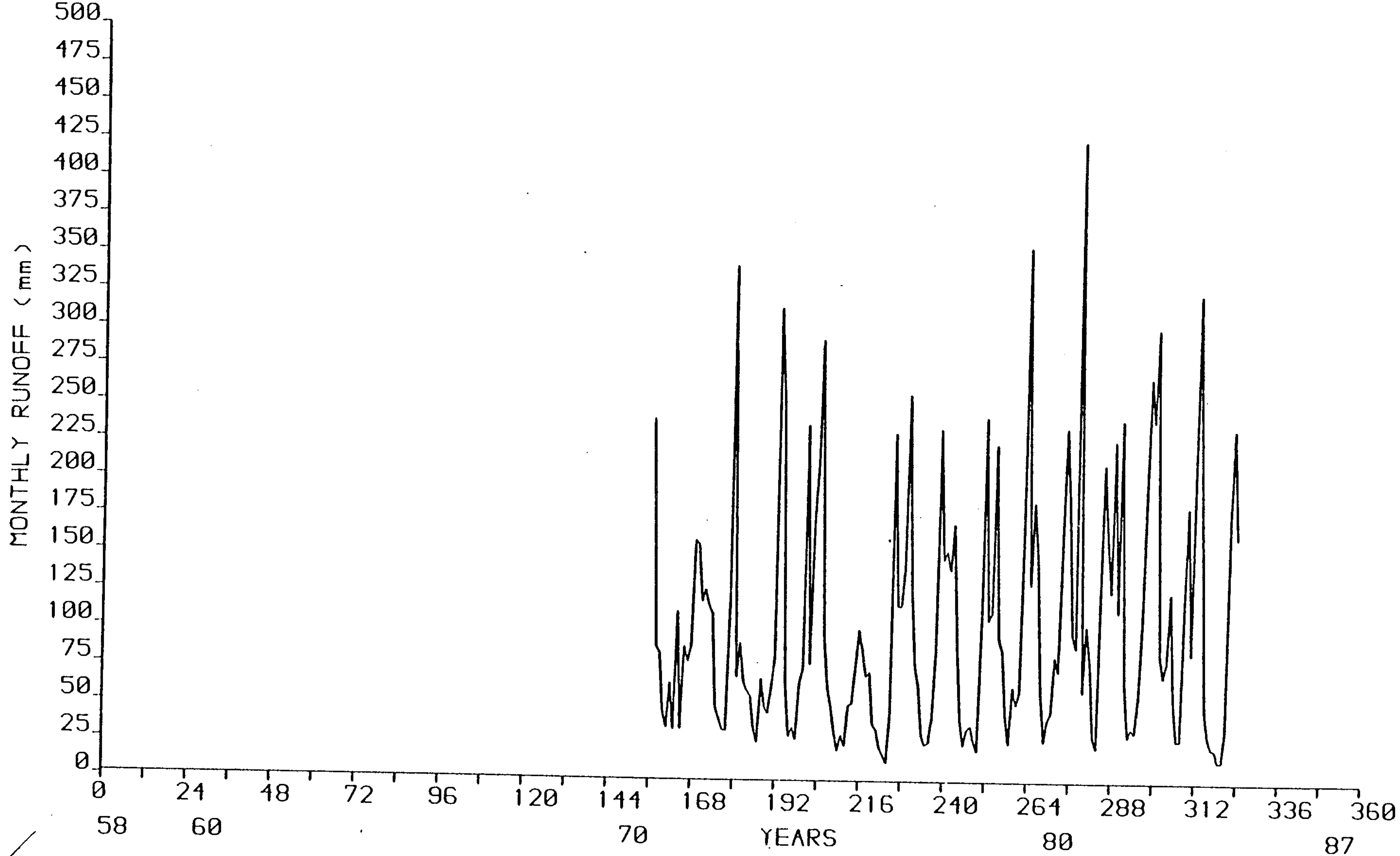


FIG. MONTHLY RUNOFF, RIVER TAFF

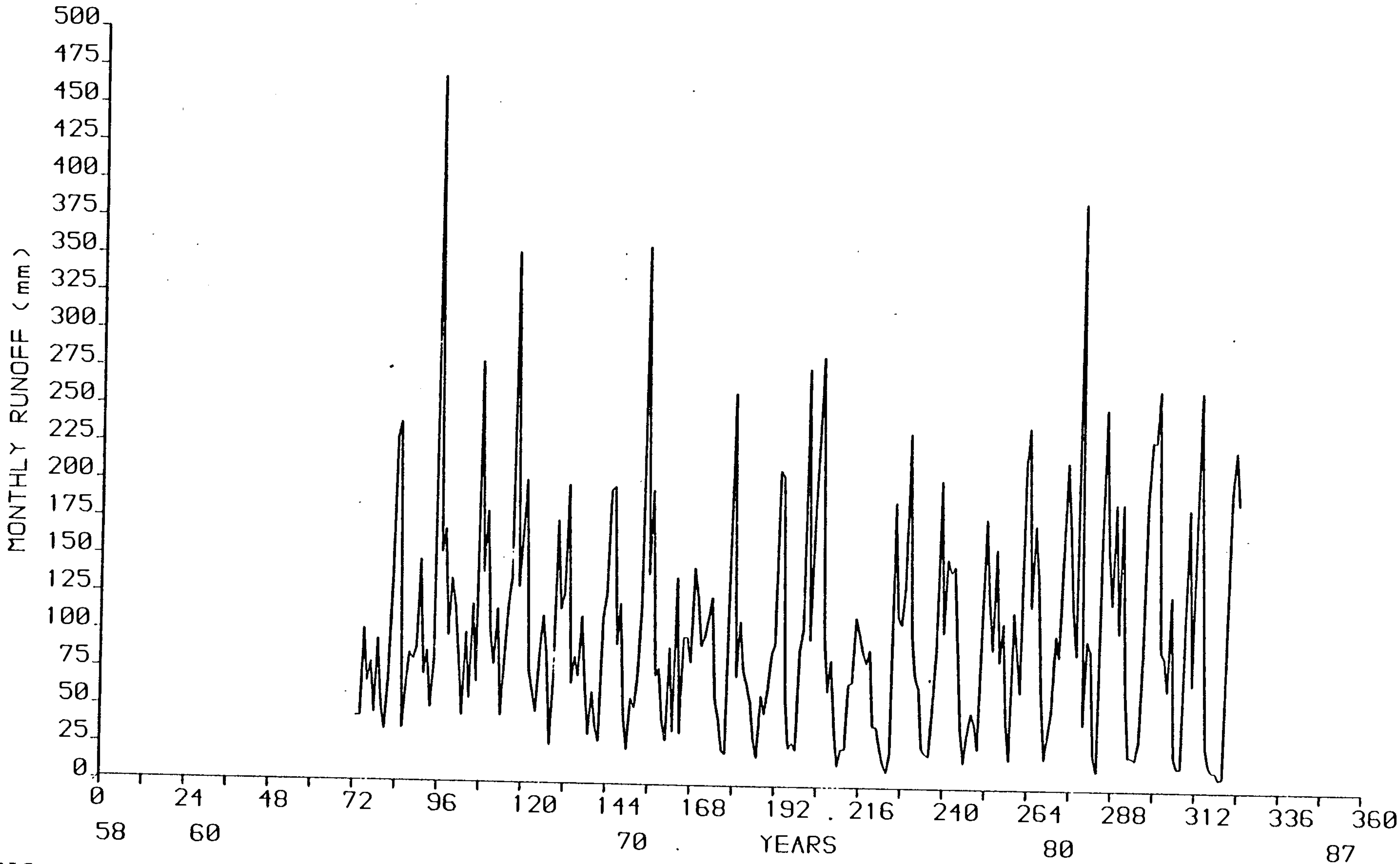


FIG. MONTHLY RUNOFF, RIVER OGMORE

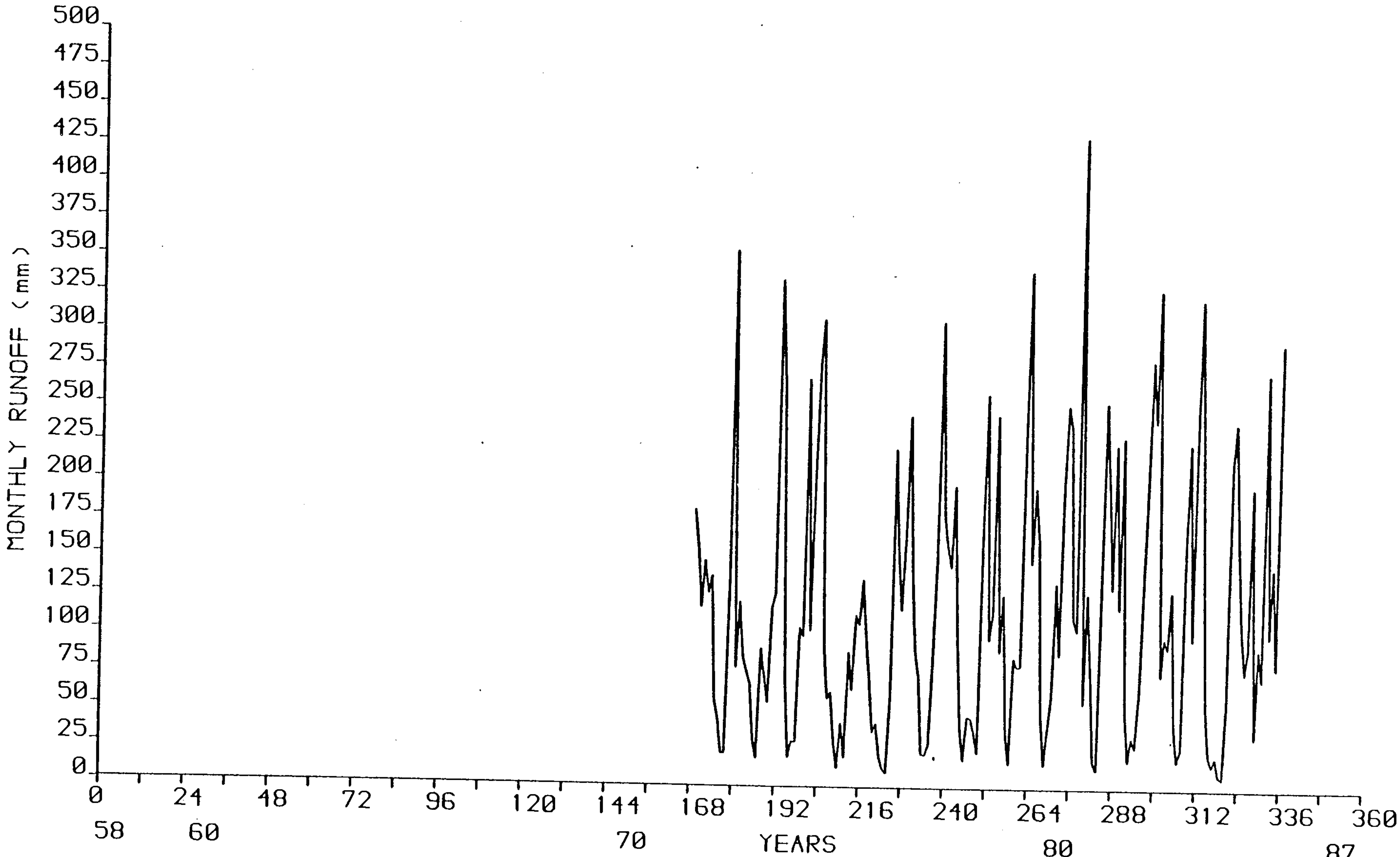


FIG. MONTHLY RUNOFF, RIVER MELLTE

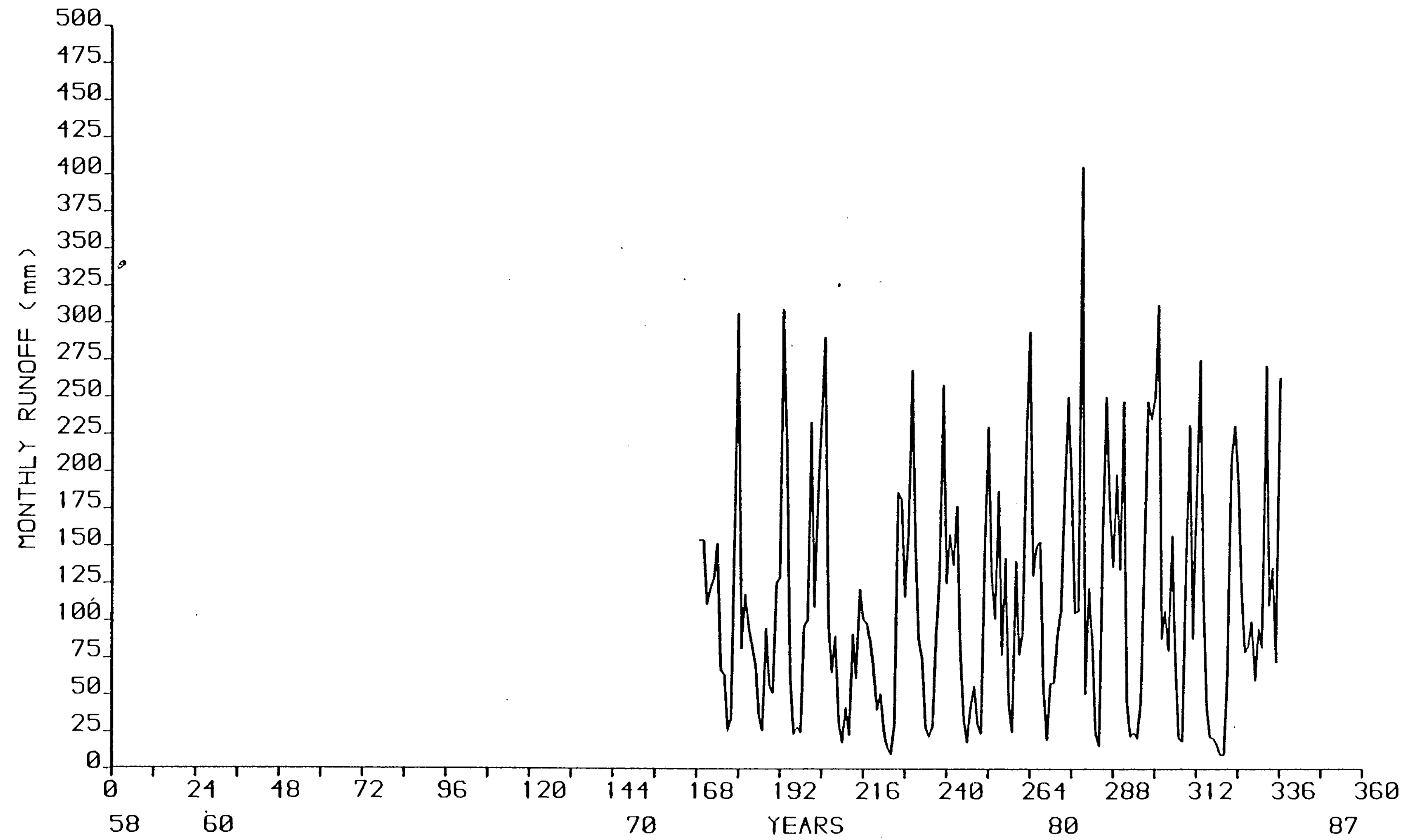


FIG. MONTHLY RUNOFF, RIVER DULAIS

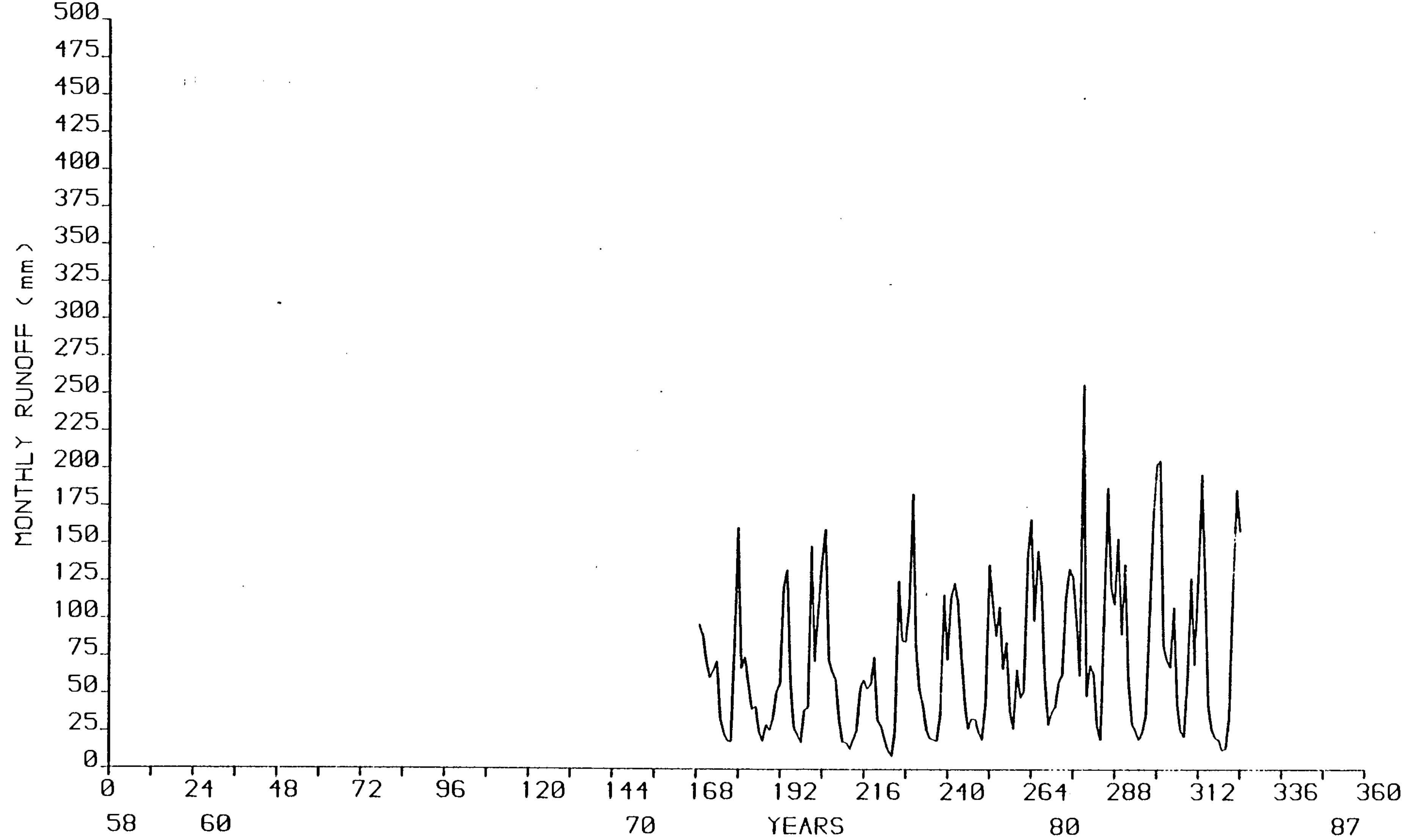


FIG. MONTHLY RUNOFF, RIVER EWENNY

Appendix 2 - Regression data

number	LARO	LARF	-a	b	Qbar/A	ADF	Q95	MDF	bfi
1	1035	1483	532	1.06	0.455	8.030	1.512	1.046	0.59
2	749	1153	293	0.90	0.354	1.480	0.159	0.045	0.52
3	970	1485	551	1.02	0.593	16.721	2.358	0.366	0.47
4	986	1415	344	0.94	0.530	3.102	0.657	0.339	0.55
5	1179	1666	693	1.12	0.908	6.558	1.001	0.341	0.45
6	857	1454	508	0.94	0.538	2.076	0.341	0.105	0.50
7	772	1251	277	0.84	0.288	2.034	0.344	0.152	0.59
8	951	1427	376	0.93	0.644	1.921	0.183	0.075	0.47
9	1189	1766	440	0.92	0.743	4.082	0.561	0.283	0.42
10	1245	1822	464	0.94	0.629	18.368	3.546	1.699	0.48
11	1238	1739	560	1.03	0.690	5.850	1.504	0.471	0.49
12	*	*	*	*	0.958	9.213	1.605	0.278	0.34
13	1433	2061	240	0.81	1.095	2.831	0.340	0.196	0.35
14	1356	1759	300	0.94	1.048	1.759	0.255	0.127	0.39
15	869	1328	528	1.05	0.382	1.583	0.320	0.163	0.58

FLOW CHARACTERISTICS

ROW	water	forest	thin	grass	crops	urb+	allgrass
1	0.001700	0.2588	0.0821	0.4290	0.1069	0.121400	0.5111
2	0.000001	0.0693	0.0679	0.2577	0.6051	0.000002	0.3256
3	0.003900	0.1203	0.1016	0.3465	0.4251	0.002601	0.4481
4	0.000001	0.1643	0.1215	0.3909	0.1585	0.164700	0.5124
5	0.005700	0.1305	0.1438	0.3379	0.3805	0.001601	0.4817
6	0.000300	0.1856	0.1154	0.3822	0.2097	0.106900	0.4976
7	0.000300	0.3184	0.0488	0.4190	0.2136	0.000002	0.4678
8	0.000001	0.0559	0.1217	0.4776	0.3445	0.000401	0.5993
9	0.000200	0.2581	0.1663	0.3800	0.0881	0.107300	0.5463
10	0.004300	0.2065	0.1816	0.4001	0.1035	0.104000	0.5817
11	0.000001	0.2138	0.1568	0.2689	0.2876	0.072900	0.4257
12	0.001500	0.2775	0.2155	0.3940	0.0501	0.061400	0.6095
13	0.002000	0.0674	0.2795	0.5847	0.0638	0.002601	0.8642
14	0.000900	0.2482	0.1401	0.4133	0.0853	0.112200	0.5534
15	0.000001	0.0710	0.1536	0.1613	0.5528	0.061200	0.3149

PROPORTIONS OF LAND COVER (1977 CLASSIFICATION)

ROW	water84	forest84	crop84	thin84	greeng84	urban84	allgr84
1	0.00210	0.2898	0.1684	0.0127	0.4005	0.1265	0.4132
2	0.00001	0.0734	0.6300	0.0267	0.2692	0.0007	0.2959
3	0.00420	0.1338	0.4530	0.0634	0.3416	0.0039	0.4050
4	0.00001	0.2575	0.1985	0.0088	0.3797	0.1553	0.3885
5	0.00670	0.1291	0.4175	0.1085	0.3348	0.0033	0.4433
6	0.00110	0.1925	0.2668	0.0235	0.4133	0.1028	0.4368
7	0.00050	0.2593	0.2277	0.0335	0.4786	0.0003	0.5121
8	0.00001	0.0623	0.4064	0.0523	0.4783	0.0007	0.5306
9	0.00040	0.2614	0.1401	0.0916	0.3621	0.1443	0.4537
10	0.00430	0.2255	0.1480	0.1114	0.3786	0.1321	0.4900
11	0.00001	0.2382	0.3167	0.0620	0.2858	0.0973	0.3478
12	0.00120	0.2636	0.0908	0.1852	0.3729	0.0863	0.5581
13	0.00130	0.0697	0.0885	0.2762	0.5597	0.0046	0.8359
14	0.00001	0.2491	0.1638	0.0821	0.3823	0.1227	0.4644
15	0.00001	0.1232	0.5465	0.0206	0.1938	0.1158	0.2144

PROPORTIONS OF LAND COVER (1984 CLASSIFICATION 1)

ROW	water844	frest844	thin844	green844	crop844	urb844	allgr844
1	0.00150	0.1104	0.0374	0.5438	0.1562	0.1463	0.5812
2	0.00001	0.0171	0.0211	0.3126	0.6478	0.0015	0.3337
3	0.00390	0.0533	0.0625	0.4018	0.4711	0.0067	0.4643
4	0.00001	0.0511	0.0304	0.5194	0.2161	0.1792	0.5498
5	0.00580	0.0583	0.1088	0.3804	0.4385	0.0066	0.4892
6	0.00050	0.0549	0.0558	0.4784	0.2802	0.1238	0.5342
7	0.00030	0.1095	0.0447	0.6192	0.2242	0.0019	0.6639
8	0.00001	0.0146	0.0281	0.5302	0.4256	0.0014	0.5583
9	0.00040	0.1099	0.1530	0.4063	0.1546	0.1658	0.5593
10	0.00380	0.1053	0.1728	0.3939	0.1568	0.1555	0.5667
11	0.00010	0.1148	0.0967	0.3066	0.3565	0.1229	0.4033
12	0.00100	0.1697	0.2374	0.3838	0.0998	0.0828	0.6212
13	0.00100	0.0345	0.3200	0.5340	0.0993	0.0102	0.8540
14	0.00010	0.1305	0.1300	0.3914	0.1984	0.1332	0.5214
15	0.00001	0.0119	0.0413	0.2071	0.6025	0.1358	0.2484

PROPORTIONS OF LAND COVER (1984 CLASSIFICATION 2)

ROW	wet77	veg77	bio77	b777	pc177
1	115.0	115.0	126.6	88.2	104.4
2	103.0	101.8	155.5	133.3	164.7
3	106.4	105.7	146.7	121.6	149.1
4	115.6	116.8	123.2	93.3	111.9
5	107.9	107.5	142.8	118.6	145.2
6	113.1	113.2	130.7	97.3	116.5
7	107.0	106.4	144.6	101.1	120.7
8	106.5	108.0	140.6	123.2	151.2
9	116.0	117.3	122.0	90.7	108.2
10	116.5	118.8	119.6	91.9	110.0
11	110.5	109.1	140.0	107.5	129.8
12	116.4	118.9	120.1	89.8	106.6
13	114.7	120.8	113.9	100.1	120.8
14	116.3	115.9	125.4	91.3	108.0
15	107.3	104.7	150.7	124.6	152.4

SPECTAL PROPERTIES 1977 IMAGE

ROW	height	S1085	STMFRQ	SOIL	LAKE	URBAN	RSMD
1	298	10.49	1.08	0.458	0.000	0.055	58.3
2	314	9.02	1.01	0.344	0.000	0.000	52.5
3	496	4.46	1.26	0.379	0.079	0.002	59.1
4	315	13.36	1.17	0.387	0.000	0.109	62.7
5	493	7.87	1.67	0.322	0.110	0.000	62.9
6	313	10.65	1.81	0.437	0.068	0.116	58.2
7	452	21.71	2.02	0.345	0.062	0.000	54.9
8	313	11.60	1.84	0.367	0.000	0.000	53.2
9	288	7.30	2.33	0.495	0.000	0.041	79.5
10	474	9.23	2.17	0.492	0.151	0.050	74.9
11	292	10.33	2.63	0.469	0.000	0.036	56.6
12	397	13.50	2.59	0.499	0.041	0.000	80.2
13	414	21.33	2.17	0.500	0.119	0.000	80.4
14	268	10.24	3.09	0.450	0.000	0.019	98.5
15	158	9.25	1.41	0.341	0.000	0.040	42.6

GEOMORPHIC PARAMETERS (FROM FSR)

Appendix 3 - Classification Procedure

SEQUENCE OF OPERATIONS TO MAKE A LAND-USE CLASSIFICATION USING THE 'ERDAS' IMAGE PROCESSING SYSTEM (Rev. 7.3)

Stage 1

Prepare the appropriate satellite image, creating the best enhancement, and rectify it to the National Grid.

Assemble documentary evidence to be used in identifying land cover types i.e. maps, aerial photogrpahy, field studies.

Stage 2

1. READ — Display the image at the high resolution monitor.
2. FIELD — Using the cursor, outline the first training areas with a polygon. The area chosen should be uniformly one type of land cover such as forest, and will thus be small, to avoid mixing with any other type of cover. FIELD then produces 2 output files, containing the statistical information for the training sample and the coordinates of the polygon verteces. The chosen name of the training sample is added to a third file.
3. ALARM — This function highlights all other areas in the image with the same statistical properties (signature) as the training field. It provides a quick but rough check on the quality of the training sample.

4. ELLIPSE — In this function, all the points of the image are plotted in a series of feature spaces, in which values in various pairs of spectral bands are displayed. Ellipses are then drawn around the area representing each training sample. This provides a powerful method of checking the training data for spurious values, and judging whether the training samples actually do separate the land cover classes or whether the training fields overlap. Judicious use of ELLIPSE will indicate those training samples which are useful, and those which can be brought together into more general classes e.g. all species of trees combined into one 'forest' class.

5. ADDSIG — This provides the facility for adding the data from a number of classes to form a new class.

6. APDSIG — New classes formed in Step 5 may be isolated, and examined individually or in relation to other general classes.

7. Steps 2 to 6 are repeated, gradually refining the signature until it is judged on the basis of ELLIPSE that the group of training areas selected is representative of the land cover types of interest, and is likely to give a clear distinction between classes.

8.	MAXCLAS	Data from the image file is classified into the chosen land cover types, and a GIS file is created containing information of land cover type only. The classification procedure is chosen and controlled by the operator. (It is useful to classify a small area of the image to check and optimise the classification parameters before proceeding with the whole image).
9.	DISPLAY	The resulting GIS file is displayed and assessed. Alternatively CLASOVE may be used to display the GIS file as an overlay to the original image file.
10.	THRESH	The operator may screen out points below a specified probability of classification, thus improving the reliability of the GIS file.
11.	CMATRIX	Produces a confusion matrix which gives further information on the reliability of the classification.
12.	SIEVE	Clumps of particular class values smaller than some specified size may be eliminated.
13.	RECODE	Individual classes may be assigned new class values. Thus, if it had appeared necessary to keep conifer and deciduous woodland separate during the classification procedure it would now be possible to assign them both to the more general class 'woodland'.
14.	LISTIT	A statistical summary of the GIS file may be produced.

15. CUTTER The data within some specified boundary e.g. an administrative boundary is extracted and transferred to a new GIS file. This file may then be displayed, annotated, analysed etc. Conversely, GIS files from contiguous areas may be stitched together to create a mosaic for a larger area.

Appendix 4 - List of work published during the course of this project

Visual interpretation of standard satellite images for the design of water resources schemes, TRE Chidley and RS Drayton, Hydrologic Applications of Space Technology, Int.Assn. for Hydrological Sciences Publication No. 160, 1986, 249-256.

Hydrologic modelling using satellite imagery, R.S. Drayton and T.R.E. Chidley, Proc.Inst.Conf. on Advanced technology for monitoring and processing global environmental data. CERMA/Remote Sensing Society, pp. 219-226. London, 1985.

The use of SPOT simulated imagery in hydrological mapping, T.R.E. Chidley and R.S. Drayton, International Journal of Remote Sensing, 1986, Vol.7, No. 6, 791-799.

Visual interpretation of satellite imagery for the estimation of key parameters in hydrologic models, R.S. Drayton, T.R.E. Chidley, S.W. Rees, A.J. Pritchard, Remote Sensing Applications in Geomorphology and Hydrology, Institute of British Geographers Annual Conference, Reading, 1986.

Dambo hydrology - an application of satellite remote sensing to water resource studies in the Third World, R.S. Drayton, Proc. Int.Sym. ISPRS Commission IV, Mapping from Modern Imagery, Int. archives of Photogrammetry and Remote Sensing, Vol. 26, Part 4, 1986, 125-135.

Simple classifiers of satellite data for hydrologic modelling, R.S. Drayton, T.R.E. Chidley, W.G. Collins, Proc. 7th Int. Symp. ISPRS Commission VII, Remote Sensing for Resources Development and Environmental Management. Int. archives of Photogrammetry and Remote Sensing, Vol. 26, Part 7, 1986, 709-712.

An appraisal of the visual interpretation of standard satellite imagery for water planning in the Third World, R.S. Drayton, T.R.E. Chidley and W.G. Collins, Int. Sem. on Photogrammetry and Remote Sensing for the Developing Countries, ISPRS, New Delhi, 1986.