SAMPLING STRATEGY AND ACCURACY ASSESSMENT

FOR

DIGITAL TERRAIN MODELLING

By

ZHILIN LI

A Thesis Submitted for

the Degree of Doctor of Philosophy (Ph.D.)

of the Faculty of Science at the University of Glasgow

Topographic Science, March 1990

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ACKNOWLEDGEMENTS

The author would like to express his sincere gratitude to his supervisor, **Prof. G. Petrie,** for all his help throughout the period of this study, especially his assistance in the writing up of this thesis which greatly improved the quality and final presentation of this work.

The author would also like to thank the **University of Glasgow** and the **CVCP** for their financial support during the period of his study.

Special thanks go to **Mr. A. Azizi,** his research colleague, for his numerous constructive discussions, for providing some of the test data and all his other help. Special thanks also go to **Mr. J. Shearer** for his critical comments on this thesis and for his other assistance, and to **Mr. D. Tait** for his vital help at particular times.

Thanks are also due to the following:

Prof. I. B. Thompson, head of the Department, for all his various kindnesses and help.

Prof. K. Torlegard and Dr. M. X. Li of the Dept. of Photogrammetry, Royal Institute of Technology (Sweden); **Prof. H. Ebner and Mr. W. Reinhardt** of the Dept. of Photogrammetry, Technical University of Munich (West Germany); and **Prof. F. Ackermann and Dr. M. Sigle** of the Dept. of Photogrammetry, University of Stuttgart (West Germany) for providing the photogrammetric data sets.

Dr. A. El-Niweiri for providing the Metric Camera data set.

Research Colleagues in this department for all their help in various aspects, especially Mr. A. A. Abdallah and Mr.S.N. Y. Al-Kazan.

Friends in Glasgow for their help and support, especially Dr. Z. Z. Du, Mr. W. Qian, Mr. Z. J. Wang, Dr. D. C. Jiang, Mr. C. L. Hu, Ms. K. Y. Wu, Mrs. B. Brown and Mr. J. Brown.

Staff in this Department for their assistance in various matters especially Mr. B. Methley, Mr. I. Gordon, Mr. M. Shand, Mr. I. Gerrard, Mr. J. Jowett, Mr. J. Keates and Miss B. Currie.

Staff in the Computing Centre for their assistance and advice.

His Parents in a small village in Southeast China where the author lived for eighteen years and **Prof. W. X. Liu** of the Dept. of Photogrammetry and Geology, Southwestern Jiaotong University (Sichuan, China) for their continuous encouragement.

Zhilin LI

March, 1990

PREFACE

This thesis is based on the author's research carried out during the period from October 1986 to September 1989, under the supervision of Prof. G. Petrie. It is documented in fifteen chapters. Apart from the introduction (Chapter 1) and concluding remarks (Chapter 15), the rest of this thesis can be divided into two parts, i.e. the part from Chapters 2 to 7 giving the background and the basic theoretical aspects of the research and the part from Chapters 8 to 14 covering the more experimentally orientated investigations - mainly concerned with the design and execution of the work carried out during this project and with the analysis of the experimental results.

Actually, this thesis covers a variety of topics related to both the theory and practice of digital terrain modelling and some of these discussions are even conducted in some detail. The only purpose of doing this has been to make this very specialised research topic more general, and thus to enhance the readability of this thesis.

Key Words:

Digital terrain modelling; Digital terrain models (DTM); Mathematical models; Terrain descriptors; Slope; Sampling strategy; Photogrammetric sampling; Photogrammetrically measured data; Contour maps; Cartographic digitisation; Digital contour data; Surface reconstruction; Surface continuity; DTM networks; DTM data quality; Gross error detection; Check points; Accuracy assessment; Accuracy models; Optimum sampling interval; etc.

With the rapid development of science and technology, especially computing technology, digital terrain modelling, which originated only thirty years ago, has already become a relatively important branch of topographic science. What is new today might be totally out of date in a few years time. In some years time, there will be no surprise when the readers of this thesis find that there is nothing new inside.

The author

A B S T R A C T

In this thesis, investigations into some of the problems related to three of the main concerns (i.e. accuracy, cost and efficiency) of digital terrain modelling have been carried out. Special attention has been given to two main issues - the establishment of a family of mathematical models which is comprehensive in theory and reliable in practice, and the development of a procedure for the determination of an optimum sampling interval for a DTM project with a specified accuracy requirement. Concretely, the following discussions or investigations have been carried out:-

i). First of all, a discussion of the theoretical background to digital terrain modelling has been conducted and an insight into the complex matter of digital terrain surface modelling has been obtained.

ii). Some investigations into the improvement of the quality of DTM source data have been carried out. In this respect, algorithms for gross error detection have been developed and a procedure for random noise filtering implemented.

iii). Experimental tests of the accuracy of DTMs derived from various data sources (i.e. aerial photography, space photography and existing contour maps) have been carried out. In the case of the DTMs derived from photogrammetrically measured data, the tests were designed deliberately to investigate the relationship between DTM accuracy and sampling interval, terrain slope and data pattern. In the case of DTMs derived from digital contour data, the tests were designed to investigate the relationship between DTM accuracy and contour interval, terrain slope and the characteristics of the data set.

iv). The problems related to the reliability of the DTM accuracy figures obtained from the results of the experimental tests have also been investigated. Some criteria have also been set for the accuracy, number and distribution of check points.

v). A family of mathematical models has been developed for the prediction of DTM accuracy. These models have been validated by experimental test data and evaluated from a theoretical standpoint. Some of the existing accuracy models have also been evaluated for comparison purposes.

vi). A procedure for the determination of the optimum sampling interval for a DTM project with a specified accuracy requirement has also been proposed. Based on this procedure, a potential sampling strategy has also been investigated.

Sampling Strategy and Accuracy Assessment
for
Digital Terrain Modelling

C O N T E N T S

Chapter 1

I N T R O D U C T I O N

Chapter One

INTRODUCTION

People live on the Earth and learn to cope with its terrain. Civil engineers design and construct buildings on it; geologists try to study its underlying construction; geomorphologists are interested in its shape and the processes by which the landscape was formed; and topographic scientists are concerned with measuring and describing its surface and presenting it in different ways, e.g. using maps, orthophotographs, etc. Despite these differences in emphasis and main interests, all of these specialists still have a common interest, i.e. they wish the surface of the terrain to be represented conveniently and with a certain accuracy. It is this widespread common interest in representing or modelling the terrain surface and the many problems associated with it that has caused the author to devote his research to this important topic.

In this introduction, first of all, the importance of carrying out a research project in digital terrain modelling in practice will be discussed; then the problems associated with digital terrain modelling will be examined; after which the research objectives of this project will be introduced. Finally, the structure of this thesis is described at the end.

1.1 The importance of carrying out research into digital terrain modelling

The importance of carrying out a research project in the field of digital terrain modelling is embodied in the important role which a digital terrain model (DTM) plays in terrain surface representaion and in related application fields.

Terrain surface can be represented in different ways. One of these could be a **painting**, and indeed this may be the oldest representation. A painting may offer some general information about the terrain which it depicts. However, the metric quality is extremely low and, in fact, it cannot be used at all for engineering purposes.

Photography is another way of recording and representing the terrain. The most useful form is aerial photography. In an aerial photograph, one dimension of the 3-D terrain, the height, is essentially absent, so that a single aerial photograph cannot be used to derive any information about the true heights of ground points. Also in metric terms, it gives information of only limited usefulness in respect of the true horizontal positions of the image recorded on the photograph.

But an **orthophotograph,** the photographic image produced after differential rectification, can be used as a plan, and indeed it may offer more information than a plan constructed using map symbols, though this information may not be available or interpretable without specialist knowledge.

A **contoured topographic map** is, perhaps, the most familiar way of representing the terrain. In a topographic map, all the features present in the terrain are projected orthogonally onto a horizontal datum. The detail is then reduced in scale and represented by lines and symbols. The terrain height and morphological information are represented selectively by contour lines.

The terrain can also be represented by a **perspective view.** The process of representing a surface in this way includes projecting it onto a plane surface and removing those lines which are not visible from the direction or point of projection. One such product is the so-called **"block diagram";** another is the **perspective contour diagram.** A digital model of the terrain surface is essential for the easy production of such representations, though laborious manual methods have been used in the past.

The real world is three-dimensional. Any system of projection on to a two-dimensional surface such as a paper sheet (or map) 'loses' one dimension. So some alternative kind of **3-D representation** is very desirable. As stated before, a digital terrain model is necessary for the generation of such a 3-D product. More importantly, in contour map production and other related areas, such a model plays an ever more important role. Therefore, carrying out some investigations into the theory and practice of digital terrain modelling is of fundamental importance for a wide range of geo-sciences.

1.2 The problems associated with digital terrain modelling

The concept of the DTM came into use in the late 1950s (Miller & Laflamme, 1958). Since then, it has received great attention and has developed rapidly in the fields of surveying, civil engineering, geology, mining engineering, landscape architecture, military operations, aircraft and battlefield simulation, etc. In spite of all this activity, there are still a lot of problems to be solved.

In order to understand the problems associated with digital terrain modelling, first of all, the **general process** needs to be described and understood. The overall process can be shown in a diagram such as Fig.1.1 which the present author has devised for the purpose. In this

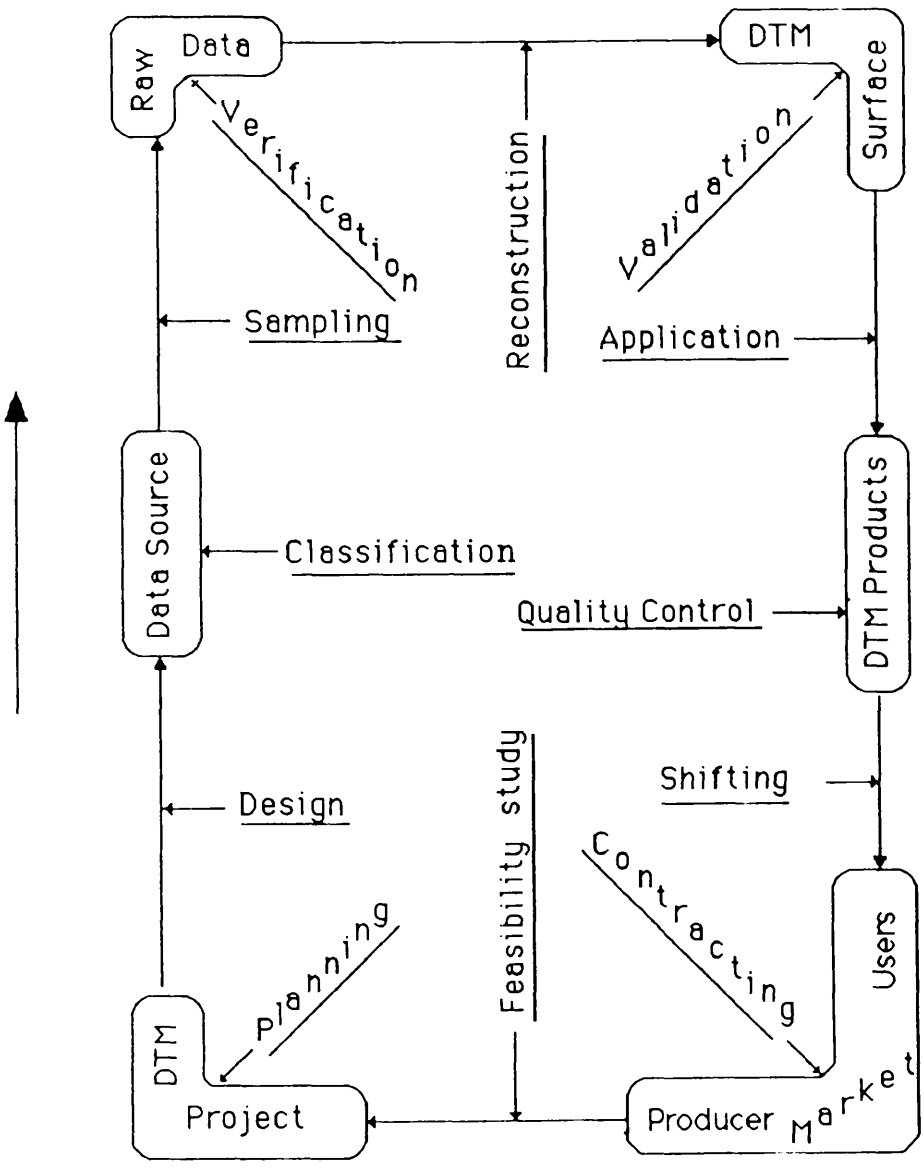


Fig.1.1 The general process of digital terrain modelling

diagram, six stages can be distinguished. Furthermore one or more operations need to be applied to the process to make it advance from one stage to another. As a result, 12 operations are distinguished. In practice, not all of these operations are necessary in the context of a particular DTM project. However, the following parts are essential, i.e. the raw data must be acquired from the data source, then the DTM surface has to be constructed from the raw data.

1.2.1 General problems associated with digital terrain modelling

For a DTM project, the final goal is to produce a DTM with a required or specified accuracy, preferably in an economical and efficient manner. That is to say, the following three criteria:-

i). accuracy;
ii). cost; and
iii). efficiency

are those of most concern to both producers and clients involved in DTM production. And these three criteria are strongly inter-related.

Among them, **accuracy** is possibly the most important single concern since usually it lies at the core of a particular DTM project. From this point of view, the accuracy of the raw data should be as high as possible; comprehensive methods and/or algorithms should be used for surface reconstruction; and the raw data should be collected as densely as possible. However, an increase in the data density will almost certainly increase the cost and possibly affect the efficiency of the modelling process. Therefore, from the viewpoints of both **economy and efficiency**, the number of data points used in the process should be kept to a minimum.

Thus, attempts to optimize both accuracy and cost are usually contradictory. In order to reach a compromise between these two factors, an **optimum density** should be used, as a result of which, the accuracy requirement can be achieved and the data points kept to a minimum. So far, this problem still remains unsolved although efforts to produce a solution have been made by several investigators. In DTM practice, the data density for a DTM project is determined more or less by experience or common practice. Obviously some more rigorous theoretical guide is desirable, especially if it can be proven or backed up by experimental results.

In order to set a rigorous theoretical guide for the determination of such an optimum density for DTM data acquisition and terrain surface representation, a **mathematical model** (or models) is necessary to predict the accuracy of the DTM (or DTM products).

From the discussions above, it can be concluded that -

i). the establishment of a family of **mathematical models for DTM accuracy prediction,** which will be comprehensive in theory and can produce a reliable prediction in practice; and

ii). the devising of procedures for the determination of **optimum data density**

are the two of the main issues in digital terrain modelling. If the former can be achieved successfully, it can be used as the basis for the latter.

1.2.2 Individual factors involved in digital terrain modelling

In the previous section, the main concerns of digital terrain modelling have been outlined and two of the main issues have been identified. Therefore, it seems pertinent next to inspect some individual factors which are associated with these general problems to see what has been done and what needs to be done.

Since it has been discussed previously that the establishment of a comprehensive mathematical model for DTM accuracy prediction is of fundamental importance, those factors which affect the accuracy of a DTM should, first of all, be identified. As one can imagine, the errors present in a DTM are the accumulated consequence of all the stages and operations involved in the terrain modelling process. Therefore, a number of factors need to be taken into account in the analysis of DTM accuracy. In particular, the following may be regarded as the main factors:-

i). the characteristics of the terrain surface;
ii). the density and distribution of the measured data, i.e. the DTM source data;
iii). the accuracy of the source data;
iv). the methods and approaches used for the DTM surface modelling and the characteristics of the finally constructed DTM surface.

The problems relating to each of these factors together with some other individual problems related to the more general problems inherent in DTMs will be examined and analysed in some detail in the following paragraphs.

(1). Terrain surface characteristics

First of all, there is the matter of the characteristics of the terrain

surface which, in the context of terrain modelling, can, to large extent, be summarized in a single concept - that of terrain **roughness.** While most people will feel intuitively that they know what is meant by this term, it is, in fact, a very abstract concept which is difficult to define exactly. Different types of descriptors have been used. Mark (1975) evaluated all the existing geomorphometric parameters. Ayeni (1976) discussed the requirements for a terrain classification system. Kubik and Botman (1976) endeavoured to use autocovariance as a descriptive parameter of the terrain. Frederiksen (1980) and his colleagues (1983, 1986), tried to use some other more complicated mathematical descriptors, e.g. the Fourier spectrum of the terrain profiles, fractal dimension and variogram. However, all of these mathematical descriptors are not too practical to use because, as Yoeli (1983) points out, "the Earth's surface, being completely irregular, is, prima facie, void of any mathematical characteristics." Therefore, what should be used as the measure of terrain roughness is still an open discussion.

(2). Density and distribution of DTM source data

In order to implement the terrain modelling process, a certain amount of data, referred to as source data in this thesis, has to be acquired. This data is then used to generate the DTM surface. The principal characteristics of the source data can be described by three parameters (referred to as attributes in this thesis), which are the **accuracy, density, and distribution** of the data. The latter two are combined in the term sampling. Thus, sampling is the process of defining the density and distribution of the points to be used for terrain modelling.

In DTM data acquisition using **photogrammetric methods,** although different sampling methods such as selective sampling, profiling, progressive sampling, composite sampling, etc. have been and are being used, sampling based on the use of a regular grid is still the most popular one. In this area, there is still no proven or accepted theory on which to base the sampling required for different types of terrain. In practice, the theoretical background to sampling is rarely discussed, thus the effectiveness of the existing sampling methods is rarely examined.

In DTM data acquisition based on existing **cartographic material,** only two methods - line-following and the raster scanning of contours - are used. At the present time, the manual line-following method dominates this area. Even here, there is no systematic theory to guide the selection of sampling intervals for different types of terrain for varying contour intervals and to satisfy the specific accuracy requirements of the DTM.

From this discussion, it can be seen that the two main parameters (density and distribution) involved in sampling should be verified or controlled. At present, the ways of doing so are limited to the optical superimposition of data points onto photographs or the injection of data points onto a stereo model (e.g. carried out in an analytical plotter) or the employment of a graphics editor (see Beerenwinkel et al, 1986; Ostman, 1986a, 1986b; Reinhardt, 1986, 1988; Uffenkamp, 1986; and Ebner and Reinhardt, 1987).

(3). Accuracy of DTM data

Considering first of all accuracy of DTM **source data**, this is related to the errors present in different stages of measurement. Different types of errors such as systematic errors, gross errors amd random errors will inevitably occur in the DTM source data Some of them need to be detected and removed from the data set and some of them lessened in extent. Up till now, not too much work has been done in this respect.

The errors likely to be encountered both in the source data and in the final DTM data should consist of two components, i.e. the horizontal error and the vertical error. So any assessment of accuracy should be applied to both of these components. **Horizontal accuracy** is not easy to assess and very little work has been done on this. Ley (1986) mentioned some methods, but he also mentioned that these are either impractical or difficult to carry out. So the question of how to assess this is still open and it is one which is difficult to answer. In contrast to the assessment of the horizontal accuracy of a DTM, more and more attention has been focused on the assessment of DTM **height accuracy.**

For a researcher working in this area, the main purpose of assessing the accuracy of a DTM is to gain experience so that some predictions may be made which may be useful for DTM practice. Such an experience may be expressed by a mathematical model, which could be either empirical or theoretical in origin. An empirical model can be obtained through experimental tests, but a theoretical model can only be obtained by a theoretical analysis.

To form an accuracy model through **experimental investigation,** an intensive test against the factors which affect DTM accuracy needs to be carried out, then some kind of empirical model can be obtained using regression techniques. However, in order to complete such a task, a huge amount of work needs to be carried out. In fact, it can be said that as yet not many comprehensive tests have been carried out. Indeed, for such a purpose, many more comprehensive experimental tests are required. This testing will be very time-consuming and almost certainly

will be very costly. Therefore, at the present time, a prior theoretical analysis is more appropriate. However, although it is impractical to establish a really satisfactory empirical model based on the results of intensive and comprehensive experimental tests, in practice, some experimental tests are unavoidable, e.g. some results are needed to validate a theoretical model.

On the **theoretical side,** quite a lot of efforts have already been made by several investigators using different techniques to establish establish satisfactory accuracy models. Usually, a numerical parameter such as those mentioned previously has been used to describe the roughness of the terrain surface. Makarovic (1972) tried to solve this problem through the analysis of a sine wave; Kubik and Botman (1976) utilized autocovariance analysis; Frederiksen (1980) used Fourier transformation to analyze the spectrum of terrain profiles; Tempfli (1980) has employed linear system analysis; Frederiksen and his colleagues (1986) also tried to use the concepts of variogram and fractal dimension. Although a lot of efforts have already been made to investigate this area, the results have not always been fruitful and there is still a lot of work to be done. In this respect, the following general remarks can be made at this stage -

i). Most of these analyses have only been applied to and concerned with measured profiles; therefore, one of the important things which needs to be done would be to extend these analyses to a complete surface.

ii). Actually, it seems to the present author that no single existing theoretical model is either convenient to use and or can be used to make a reliable prediction of DTM accuracy. Therefore, more comprehensive and reliable models need to be devised and investigated as a matter of some urgency.

Also it appears to the present author that there is no systematic theoretical basis to tell users how **check points** should be employed in DTM accuracy assessments, i.e. how many such points are essential, where they should be located and what is the required accuracy of these check points, for a given DTM accuracy requirement.

Furthermore, a lot of efforts have been made for the devising of procedures for the determination of **optimum data density,** but, the result has not been fruitful due to the lack of comprehensive and reliable accuracy models. Again, this points to the urgency and importance of establishing a family of new mathematical models of DTM accuracy.

(4). The methods used for terrain surface modelling and the the characteristics of the resulting surface

In the development of DTM methodology, the first main development has been in the field of **interpolation** methods. In this respect, someone tries to devise an interpolation algorithm; then its effectiveness is tested using one or more sets of test data; and finally there is a declaration as to how good the method is. A comprehensive review of these interpolation methods (or algorithms) has been given by Schut (1976), which is still largely valid today. Although a lot of different interpolation methods have been proposed, there is still very little detailed information and few practical or methodological comparisons between the different techniques which have actually been used in practice (Ackermann, 1979).

Furthermore, there is a lot of confusion in this area. Indeed, this is one of the most serious problems which bedevil digital terrain modelling. Actually, interpolation techniques have been applied mainly for two quite different purposes, i.e. either for the pre-processing of data such as the so-called random-to-grid interpolation before surface modelling or for the computation of DTM points from a DTM surface after surface modelling. Therefore, the discussion of interpolation should be related to the **methods and approaches** used for surface modelling. Indeed, all of this matter needs to be clarified.

Although, initially grid-based modelling has been the basic modelling approach of most DTMs, **triangulation** is another alternative to form networks for surface reconstruction which has become popular in recent years. In this respect, several triangulation algorithms or procedures have been developed by different authors, e.g. McLain (1976), Yoeli (1977), Elfick (1979), McCullagh and Ross (1980), Mirante and Weigarten (1982), etc. In addition, some contour-specific triangulation methods have also been designed by several investigators such as Christensen (1987). The main problem associated with triangulation procedures is how to accommodate distinctive terrain features in the terrain model, especially break lines and form lines. It seems to the present author that, till now, no single triangulation procedure has solved this problem comprehensively.

From above discussion, it is obvious that any researcher trying to advance the methodology of digital terrain modelling must at all times keep the matter of the specific methods used for modelling the terrain surface in the forefront of his mind since they bear so heavily on the characteristics of the particular surface which will result.

1.3 Research objectives

As discussed above, accuracy, economics and efficiency are three of the most important concerns in digital terrain modelling. To carry out some work related to these three concerns is, of course, the objective of this project. However, many problems related to each of these concerns have already been pointed out in the previous section. Therefore, no attempt can be made or has been made to solve all of these problems due to the limited period of time available, and the basic unsolvability of some parts of those problems - at least at the present time. However, to solve or to contribute to a solution of some of the problems related to digital terrain modelling is the **main objective** of this research project.

In more specific terms, research into the following areas will be carried out in the course of this project:

i). First of all, **photogrammetric sampling** will be examined from various viewpoints; existing sampling methods will also be scrutinised; and the problem of optimizing regular grid sampling will also be considered.

ii). The **errors** which occur in **source data** will also be under investigation. Attempts will be made to develop some algorithms for gross error detection and random noise filtering. These algorithms will also be validated experimentally.

iii). Some **experimental tests** of the accuracy of DTMs derived from the data sets acquired from different data sources (i.e. aerial photographs, space photographs, and existing contour maps) with different data attributes will be carried out. In the case of DTMs derived from photogrammetrically measured data sets, special attention will be paid to the **variation of DTM accuracy with data density** (expressed in terms of sampling interval).

iv). In the context of experimental tests of DTM accuracy, an attempt will be made to carry out a theoretical analysis of the effects of the characteristics of **check points** on the reliability of the DTM accuracy figures which have been derived from the test results and then to set some theoretical guidelines for the requirements of check points, i.e. the minimum number needed and their required accuracy, as well as the desirable distribution of such points.

v). A family of **mathematical models** will be established for the height accuracies of DTMs derived from photogrammetrically measured elevation data. These models will be validated using

experimental test results and evaluated using some theoretical standards. Some existing accuracy models will also be evaluated using test data generated by this project.

vi). After the establishment of these models, some procedures for determining the **optimum density** of DTM source data in different cases will be proposed.

1.4 The structure of this thesis

Apart from this introductory chapter, the rest of this thesis is organised as follows:

Chapter 2 discusses some of the theoretical background to modelling in general and digital terrain modelling in particular. This is followed in **Chapter 3** by a discussion of the main terrain descriptors used in the geo-sciences. In this chapter, not only are the different types of terrain descriptors discussed but the adequacy of using slope as the principal terrain descriptor for terrain modelling purposes is also emphasized.

Chapter 4 is concerned with photogrammetric sampling strategy. The theoretical background to photogrammetric sampling and the requirements of different sampling methods are discussed and analysed. Photogrammetric sampling is also examined from various other points of view such as those employed in statistics, geometry, geomorphology and topographic science; existing sampling methods are briefly reviewed; and a critical analysis of the most important attributes of photogrammetric sampling will also be undertaken.

Chapter 5 discusses briefly the accuracy of photogrammetrically measured DTM source data. The acquisition of DTM source data from existing maps is reviewed in **Chapter 6**, where the vexed question of the estimation of the accuracy of digitised contour data is also discussed.

In **Chapter 7**, the different approaches to the construction of DTM surfaces are outlined; some discussion of the characteristics of DTM surfaces is also given; existing knowledge in this field is reviewed critically.

Chapter 8 is allocated to a discussion of possible methods for an improvement in the quality of measured DTM source data. Algorithms for detecting gross error in source data sets have been devised and are presented and tested experimentally. The effect of random noise on the quality of raw data is also discussed and has been investigated experimentally.

In **Chapter 9,** various alternative plans for carrying out experimental tests on DTM accuracy are outlined. This is followed by a description of a specially designed experimental test of the variation of DTM accuracy with the density (expressed in terms of sampling interval) of DTM source data. Finally, a critical analysis of the results of this test is conducted.

As an extension of Chapter 9, an experimental test on the accuracy of DTMs derived from space photography taken by the Metric Camera will be described in **Chapter 10.**

Chapter 11 describes the theory behind the use of check points in experimental tests of DTM accuracy. Three problems (attributes) are discussed, i.e. the required accuracy, the minimum number of points required and the desirable distribution of these points. The relationship between these and the reliability of DTM accuracy estimates is also established. This theoretical work has also been validated with data from the experimental tests.

Chapter 12 is concerned with the theoretical aspects of DTM accuracy assessment. Existing mathematical models of DTM accuracy will be evaluated both experimentally and theoretically. A new family of mathematical models has also been established. These models will be evaluated from a theoretical viewpoint and validated experimentally.

The discussions in the previous chapters (from 8 to 12) are more or less related to the accuracy of DTMs derived from photogrammetrically measured data. In **Chapter 13,** two sets of experimental tests into the accuracy of the DTMs derived from contour maps for different purposes will also be described.

In **Chapter 14,** some procedures for the determination of optimum data density will be presented. In this chapter, a discussion of the optimization of regular grid sampling will also be carried out.

In **Chapter 15,** some concluding remarks on the results of this research project are given, together with recommendations for future research.

Following this general introduction, the next chapter will contain a more detailed discussion of the theoretical background to digital terrain modelling.

Chapter 2

Mathematical Models and Digital Terrain Modelling

Chapter Two

Mathematical Models and Digital Terrain Modelling

In the previous introductory chapter, four topics have been discussed, i.e. the importance of undertaking a research project in digital terrain modelling; the main problems associated with digital terrain modelling; the objectives of this research project; and finally the structure of this thesis. In this chapter, a more detailed introduction to digital terrain modelling will be given to act as background to the discussions about this subject which will take place later in this thesis. More specifically, general concepts about models and mathematical models are introduced and reviewed; concepts and technical terms used with digital terrain models are introduced and discussed; and finally a brief discussion on the mathematical models used in digital terrain modelling is also given.

2.1 About models in general

It seems pertinent first to have a brief discussion about the general concept of a model. A typical definition of the term is the following:

> "**A model** is an object or a concept that is used to represent something else. It is reality scaled down and converted to a form which we can comprehend" (Meyer, 1985).

A model may have a few specific purposes such as prediction, control, etc., in which case, the model only needs to have just enough significant detail to satisfy these purposes. The model may be used to represent the original situation (system or phenomenon) or it may be used to represent some proposed or predicted situation.

Thus, the word "**model**" usually means a representation and, in many situations, it is used to describe the system at hand. Consequently, there are strong differences of opinion as to the appropriate use of the word "model". For example, it may be applied to a photogrammetric replication of a piece of the terrain surface which has been photographed or it may suggest a perspective view of the piece of the terrain.

Generally speaking, there are three kinds of models as follows:

i). conceptual;
ii). physical (analogue); and
iii). mathematical models.

The **Conceptual model** is the model borne in a person's mind about a

situation or an object based on his knowledge or experience. Often, this particular type of model forms the primary stage of modelling and will be followed later by a physical or mathematical model. However, if the situation or object is too difficult to represent in any other way, then the modelling will remain in conceptual form.

A **Physical model** is usually an analogue model. An example of this kind of model would be a terrain model made of rubber, plastic, or clay. A stereo-model of the terrain based on optical or mechanical projection principles, which has a widespread use in photogrammetry, would also fall into this category. The size of a physical model is usually much smaller than the real object.

A **Mathematical model** is the representation of a situation, an object, or a phenomenon in mathematical terms. In other words, a mathematical model is a model whose component parts are mathematical concepts, such as constants, variables, functions, equations, inequalities, etc. In the context of stereo-photogrammetry, the most obvious examples of this type of model are those used in analytical photogrammetry based on the use of collinearity or coplanarity equations.

2.2 About mathematical models

General speaking, mathematical models may be divided into two types (Saaty and Alexander, 1981) as follows:-

i). quantitative; and
ii). qualitative.

Quantitative models are based on the number system, while **qualitative** models are, on the other hand, possibly based on set theory, not reducible to numbers.

Also, a problem may either be **deterministic** or be subject to changes and therefore **probabilistic.** Therefore, mathematical models may be classified into:-

i). **functional** models, which are those intended to solve deterministic problems, and

ii). **stochastic** models, which are those used to solve probabilistic problems.

2.2.1 Advantages of mathematical models

One very important question about mathematical models is "what kind of

benefit can one have by using mathematical models" or "why should we make use of mathematical models" ? Saaty and Alexander (1981) give some reasons for using mathematical models as follows:

i). Models permit abstractions based on logical formulation using a convenient language expressed in a shorthand notation, thus enabling one to visualise better the main elements of a problem while, at the same time, satisfying communication, decreasing ambiguity, and improving the chance of agreement on the results;

ii). A model allows one to keep track of a line of thought, focusing attention on the important parts of the problem;

iii). Models help one to generalise or to apply the results of solving problems in other areas;

iv). They also provide an opportunity to consider all the possibilities, to evaluate alternatives, and to eliminate the impossible ones; and

v). They are tools for understanding the real world and discovering natural laws.

2.2.2 Standards for evaluating mathematical models

From the foregoing discussion, it appears that some benefits can be gained by using mathematical models. Now comes the question -"what kind of mathematical models should be used" ? This is related to the problem of "how to judge the 'goodness' or value of a mathematical model".

As one can imagine, there must be certain characteristics which all models must have to a varying degree and which bear on the question of how good they are. These characteristics can be used as standards for model evaluation. Meyer (1985) suggests the following as being important:

i). accuracy;
ii). descriptive realism;
iii). precision;
iv). robustness;
v). generality; and
vi). fruitfulness.

A model is said to be **accurate** if the output of the model (i.e. the answer it gives) is correct or very near to correct. A model is said to be **descriptively realistic** if it is based on assumptions which are correct. A model is said to be **precise** if its predictions are definite

numbers (or other definite kinds of mathematical entities, such as functions, geometric figures, etc.). By contrast, if a model's prediction is a range of numbers (or a set of fuctions, a set of figures, etc.), the model is said to be imprecise. A model is said to be **robust** if it is relatively immune to errors in the input data. A model is said to be **general** if it applies to a wide variety of situations. A model is said to be **fruitful** if either its conclusions are useful or it inspires and/or points the way to other good models (Meyer, 1985).

2.2.3 Practical considerations for mathematical models

The six characteristics discussed above can be regarded as the theoretical standards for model evaluation. However, in modelling practice, other practical considerations should also be taken into account.

In practice, the situation may arise that two models are both good in terms of these characteristics and the question which then arises is which should be selected for practical use. In this case, it could be that one is simpler than the other and could be preferred. Thus, it can be seen that the **simplicity** (or complexity) of models could or should become a very important criterion for the comparison or selection of models.

Indeed, the basic premise in modelling is that complicated models are not always needed even though a phenomenon may be complicated. It has been stated as **the principle of parsimony** (Cryer, 1986) that the (mathematical) model used should require the smallest possible number of parameters that will adequately represent the data sampled from the phenomenon (or so-called reality).

Evidence from many fields shows that mathematical models with a degree of complexity beyond a certain level often perform poorly in comparison with other simpler models when reality is taken into account. Often, if a mathematical model for a given phenomenon involves a large number of parameters, it is a good indication that an entirely different family of models should be considered for the representation of this phenomenon.

Another situation may arise in which a model may be unable to produce a very accurate representation or prediction for a situation or phenomenon, in which case, what kind of attitude should be taken towards it ? In practice, for a mathematical model, what should be considered, rather than rejecting it, is to try and estimate how wrong it is. If a model can still produce reasonable results in any case, it can still be regarded as a robust model.

2.3 About digital terrain models

2.3.1 The introduction of digital terrain models

Terrain models have always appealed to military personnel, planners, landscape architects, civil engineers, as well as specialists in the various disciplines of the earth sciences. Originally, the models were physical models, made of rubber, plastic, as well as clay, sand, etc. For example, during the Second World War, many models were made by the American Navy and reproduced in rubber (Baffisfore, 1957). In the recent Falklands War in 1982, the British forces in the field used sand and clay models extensively to plan military operations.

The introduction of mathematical, numerical and digital techniques to terrain modelling owes much to the activities of photogrammetrists working in the field of civil engineering. In the 1950s, photogrammetry had begun to be used widely to collect data for highway design. Roberts (1957) first proposed the use of a digital computer with photogrammetry as a new tool for acquiring data for planning and design in highway engineering. Miller & Laflamme (1958) of the Massachussetts Institute of Technology (MIT) described such a development in more detail, and introduced the concept of the **digital terrain model** (DTM). The definition given by them is as follows:

> "The digital terrain model (DTM) is simply a statistical representation of the continuous surface of the ground by a large number of selected points with known X,Y,Z coordinates in an arbitrary coordinate field."

In summary, DTM is simply a digital (numerical) representation of the terrain surface.

2.3.2 Technical terms regarding digital terrain modelling

Since Miller and Laflamme (1958) coined the original term (digital terrain model), some other (alternative) terms have been brought into use. These include digital elevation model (DEM), digital height model (DHM), digital ground model (DGM), as well as digital terrain elevation model (DTEM). These terms originated from different countries. DEM is widely used in America; DHM came from Germany; DGM is used in the UK; and DTEM was introduced and used by the USGS and DMA (Petrie and Kennie, 1987).

Some restrictions in their usage may be desirable. So Yoeli (1983) proposed to restrict the use of term DTM to the height points on the Earth's surface, and use the term DEM (or DHM) for all other

geographical and geological phenomena. This proposal might solve the confusion for modelling the Earth's surface, but leaves unsolved that for surfaces underwater. However, Yoeli's proposal has not been accepted by the rest of surveying and mapping community, so some further discussion of this point may be of value.

In practice, these terms (DTM, DEM, DHM and DTEM) are often assumed by many people to be synonymous and indeed this may often be the case. But sometimes they actually refer to quite distant products. That is to say, there may be slight differences between these terms. A selection of the meanings of these related words taken from dictionaries or literature includes the following:

Ground: "the solid surface of the Earth", "a solid base or foundation"; "a surface of earth"; "bottom of the sea"; etc.

Height: "measurement from base to top"; "elevation above ground or recognized level, esp. that of sea"; "distance upwards"; etc.

Elevation: "height above a given level, esp. that of the sea"; "height above the horizon"; etc.

Terrain: "tract of country considered with regard to its natural features, etc."; "an extent of ground, region, territory"; etc.

From these definitions, some of the differences between DGM, DHM, DEM and DTM begin to manifest themselves. So a DGM more or less has the meaning of "digital model of a solid surface". In contrast to the use of "ground", the terms "height" and "elevation" emphasize the "measurement from a datum to the top" of an object. They do not necessarily refer to the altitude of the terrain surface, but in practice, this is the aspect which is emphasized in the use of terms DHM and DEM. The meaning of "terrain" is more complex and embracing. It may contain the concept of "height" (or "elevation"), but it attempts to include other geographic elements and natural features. Therefore, the term DTM tends to have a wider meaning than DHM or DEM and will attempt to incorporate specific terrain features such as rivers, ridge lines, break lines, etc. into the model.

2.4 Mathematical models of the terrain surface

Digital terrain modelling is a process of mathematical modelling. And a digital terrain model was defined as a digital representation of

the terrain surface. In digital terrain modelling, a sample of the points forming the surface of the terrain under concern is measured with a certain accuracy. In this sense, the terrain surface is represented by a set of digital (or numerical) data. When information about the height values of other points present on the surface of this particular piece of terrain is needed, then an **interpolation** process is applied to the digital data set. In interpolation, a mathematical model is selected and used for the construction of a model of the terrain surface on the basis of the measured digital data points. The height value of any other point lying on this surface can then be obtained from this mathematical model.

A variety of mathematical models have been in use for this purpose such as a global polynomial surface, a linked series of local surfaces, a contiguous set of linear facets, etc. The general expression can be written as follows:

$$H = f(X,Y) \qquad (2.1)$$

This simply means that H, the height value of a point on the surface whose location is defined by its X and Y coordinates is a function of two variables, which are the planimetric coordinates defining its position on the surface.

Expression (2.1) can also be rewritten in another form as follows:

$$f(X,Y,H) = 0 \qquad (2.2)$$

This definition will suffice for this introduction; a more detailed discussion of this topic will be given in Chapter 7.

2.5 Mathematical models of DTM accuracy

It has been mentioned above that accuracy is an important criterion or standard for the evaluation of models in general and that the accuracy of the mathematical model used to represent the corresponding terrain surface is a matter of the highest concern in digital terrain modelling. Therefore, a mathematical model for the assessment or prediction of DTM accuracy is of importance both in theory and in practice.

The model for assessing or predicting DTM accuracy will be more complicated than that for the terrain surface itself because the latter has only two variables (X and Y coordinates) while the former will have quite a few other variables. These variables may include the roughness of the terrain surface, the specific interpolation functions and methods used, as well as the accuracy, density and distribution of the source

data, and so on. Therefore, the mathematical models of DTM accuracy could be written in a general form as follows:

$$Ac(DTM) = f\ (S, M, R, A, Ds, Dn, O) \qquad (2.3)$$

Where, Ac(DTM) denotes the accuracy of a DTM;
S is a factor related to the characteristics of DTM surface;
M is the method used for surface modelling;
R is the roughness of the terrain surface;
A, Ds and Dn are the accuracy, distribution and density of source data, respectively; and
O refers to other factors.

Also, as will be discussed later in Chapter 12, there are many accuracy models in use which take different forms and which employ different mathematical tools. Even so, it is still one of the most important topics in digital terrain modelling. Indeed, the building of this type of accuracy model in digital terrain modelling occupies a major part of this project.

As stated in the introduction, the model expressed by (2.3) may be established either by theoretical analysis or experimental tests or a combination of both. In experimental tests of DTM accuracy, check points will be used. The relationship between the reliability of the obtained accuracy results with the attributes (number, accuracy and distribution) of the check points may also be expressed in a mathematical form, thus a mathematical model may be formed. Such a model may also be expressed by an equation similar to (2.3) as follows:

$$R = f(A, No, Ds) \qquad (2.4)$$

Where, R refers to the reliability of the obtained accuracy estimates;
A, No, and Ds denotes the accuracy, the number and the distribution of the check points used for the experimental test.

2.6 Other mathematical models in digital terrain modelling

These two types of mathematical models discussed above are the main models in use in digital terrain modelling. However, if it is viewed as an engineering system, another two types of main models may be considered. One of these models is the **economic model** to be applied in digital terrain modelling. It is not a easy job to establish such a model since all the costs involved in different stages of the modelling process should be taken into consideration. In doing so, both theoretical data and practical (or empirical) values or data are also needed if the model is to be effective. It can be expected that this

type of model can only be established if DTM production has become a routine operation.

Another type of model which can be considered within the context of digital terrain modelling is the **optimization model.** This type of model can be used to design the optimum combination of data sources, instrumentation, sampling strategy, interpolation method, and other factors involved in the operation. The model should take both economic factors and accuracy into consideration or only one of them. Mathematical tools such as dynamic programming and/or linear programming may be used for the establishment of such models.

Of course, there could be quite a lot of other minor mathematical models used in digital terrain modelling; however, it is not the purpose of this section to examine all these minor models.

2.7 Simulation

Before ending this chapter, it seems pertinent to have an examination of the term, **simulation,** because simulation is relevant to modelling.

First of all, it would be appropriate to have a look at the position of simulation in the manipulations of mathematical models. It is commonplace to say that there are two great pillars upon which the experimental sciences rest - theory and experiment. Similarly, as Meyer (1985) pointed out, in mathematical manipulation, there are also two approaches. One is the **analytical approach,** which used to be and still is regarded as being the greater pillar; the other might almost be called the **experimental approach.** Simulation is one form of mathematical experiment. The difference between an analytical solution and a solution determined through a simulation is that the former involves finding a formula that relates the quantiity which we are trying to estimate to other quantities known to us, thus to provide a closed-form expression in terms of defining parameters; while the latter attempts to estimate the value of a quantity by mimicking (simulating) the dynamic behaviour of the system involved, by employing a procedural model.

As a matter of fact, in operational research, a simulation is defined as "a model of some situation in which the elements of the situation are represented by **arithmetic and logical processes** that can be executed on a computer to predict the dynamic properties of the situation" (Emshoff & Sisson, 1970). This emphasis on **dynamic behaviour** or properties is or appears to be a fundamental characteristic of simulation.

Simulation was initially developed as a way of tackling problems for which no explicit mathematical solution can be devised. However, as the method has developed, it has been seen to be very powerful, thus it is widely used in practice. Rivett (1972) has stated that there are in general three reasons for using simulation as follows:

i). Technical problems may be too complex so that an analytical approach breaks down. As stated above, this is the situation, for which simulation was initially developed. For example, to produce a mathematical model for the overall effect of different error sources on stereo-compilation simultaneously is so complex that Alspaugh (1985) made use of simulated images to form stereo-models in his experimental study (on the effects of different error sources on dynamic profiling) for the purpose of isolating the effects of some errors from others. Those images are simulated by using a digital terrain model and the so-called surface shadowing technique so that some parameters such as orientation elements and heights of image points can be known beforehand.

ii). Researchers or users need to gain some understanding of a complex real situation. An example of such a situation related to a digital terrain model is the well-known device called an aircraft **simulator**, which always includes a digital terrain model to impart some sense of reality to the trainee or pilot using the simulator.

iii). Researchers may be dealing with problems which do not as yet exist in the real world. For example, before the SPOT satellite was launched, the characteristics of SPOT images have been studied by using simulated SPOT images which could be used to form stereo-models, etc for the production of digital terrain data.

2.8 Concluding remarks

In this chapter, the concepts of a model and modelling have first been introduced and reviewed; then some discussions of mathematical models have been made - the concepts were given; their advantages were reviewed; and the standards for model evaluation have also been examined. This general introduction has been followed by the definitions and discussions of digital terrain models; after which, the applicability of mathematical models in digital terrain modelling has been discussed. Finally, a short discussion of simulation has been included with a reference to its relationship to modelling in general and to terrain modelling in particular.

The purpose of including such an introduction is to provide some background for more detailed discussions which will be included later in this thesis. Thus, an understanding of the basic concepts of models, mathematical models and digital terrain models will be helpful in understanding what digital terrain models are; and how they may be applied. The standards for model evaluation could be used later as the basis for evaluating both existing mathematical models for DTM accuracy and those which have been developed in this project. Also the discussions on the practical considerations of mathematical models will be useful in comparing these models.

After these definitions and introductory discussions about models and terrain models, it is time to consider the main characteristics of the object being modelled, that of the terrain itself. The discussions will start with the matter of terrain descriptors and terrain classification, which form the principal topics of the next chapter.

Chapter 3

Terrain Descriptors

Chapter Three

Terrain Descriptors

In order to determine the optimum sampling density of data points and to select and/or design the best function for surface reconstruction, as well as to estimate the accuracy of a digital terrain model, it is necessary to have some knowledge about the actual terrain surface.

The terrain surface can be described in different ways. In this chapter, first of all, a general discussion will take place regarding the different terrain descriptors which may be used by different specialists. This will be followed by a more detailed discussion of those descriptors which are of special importance in digital terrain modelling.

3.1 Introduction

To describe the terrain surface is a necessary preliminary to understanding its myriad characteristics. As stated in the introduction, the terrain surface is a matter of concern to specialists in a wide variety of disciplines. Those disciplines such as geology, hydrology, geography, botany, zoology, ecology, pedology and meteorology can be considered as being mainly **scientific** in nature, while others such as agriculture, forestry, civil engineering, military engineering, urban and rural landscape design, and mapping can be regarded as being mainly **applied.** Obviously, many of these fields overlap so that no rigid distinctions should be made.

The **specialists in the academic earth sciences** (including geologists, pedologists, geomorphologists, zoologists, botanists, ecologists, etc.) may be as much concerned with past processes as with future possibilities. The **specialists in agriculture** (including foresters, agriculturists, pastoralists, planters, etc.) are mainly concerned with three properties of the land apart from its location: those of soil fertility, soil manageability and the nature of the existing vegetation. **Civil engineers** consider the suitability of ground for bearing, compressibility and shear strength of soil under different moisture conditions. Usually these considerations are made in the context of the design and construction of structures such as roads, dams, buildings, railways, harbours, open-cast pits, etc. in which earthwork quantities will often be a matter of major importance. **Military engineers** may focus on less permanent works, but also on such aspects as artillery lines of sight, the suitability of ground for excavating trenches, fortifications, holding tent pegs, laying minefields, accepting parachute drops, and sustaining the repeated

passage of troops and both tracked and untracked vehicles. The **specialists in meteorology and climatology** are concerned with the effect of terrain on weather and climate because absolute elevation, slope, aspect, and the nature of the soil surface all influence climate both directly through their effect on winds, insolation, fog, cloud as well as rain and indirectly through the effects of vegetation cover. **Hydrologists** are especially concerned with runoff regimes and quantities, with stream flow, and with ground water infiltration, depths and movements with practical application to water supply, soil erosion and flood hazards, etc. (Mitchell, 1973).

The list of these interests could be very lengthy. However, it is not the purpose of this section to provide a survey of all these interests. What the author wishes to point out is that, for different types of specialists, the descriptors used for terrain classification could be very different arising from their different interests.

In general, two basic types of descriptors may be distinguished as follows:

i). **qualitative descriptors,** which are expressed in some kind of general terms, so they are referred to as **general descriptors** in this thesis; and

ii). **quantitative descriptors,** which are those specified by some kind of numerical value such as a statistical value, so they are referred to as **numerical descriptors** in this thesis.

In this chapter, the general descriptors are examined in outline only; whereas those numerical descriptors which have been used in DTM practice are discussed in rather more detail. Some comments on the suitability of these descriptors in DTM practice are also made and a specific descriptor is recommended as a critical factor to be considered in DTM work.

3.2 General (or qualitative) terrain descriptors

Based on these different interests outlined above, a variety of general or qualitative descriptors can be found. However, some are irrelevant to the concerns of topographic scientists. For example, it does not matter to topographic scientists whether or not a piece of terrain is suitable for planting or cultivation. Actually, it is only the descriptors of the roughness and the coverage of the terrain surface, especially the former, that are mainly of interest in the context of a DTM because they affect the accuracy of the photogrammetric and/or other measurement and thus the accuracy of the final DTM. Therefore,

only these related descriptors will be examined briefly in this section.

3.2.1 Terrain descriptors based on the characteristics of terrain surface cover

The terrain surface is covered with a rich content such as vegetation, soil, water, ice, artificial features and so on. These various feature which cover the terrain surface also affects the accuracy of photogrammetric measurement but the degree of this effect varies with the specific type of feature, e.g. the type of vegetation, and even with the season. The specific question about how this terrain surface cover affects the measurement accuracy is one which lies outside the main scope of this discussion. What is of interest here is the fact that, given a particular type of terrain, some idea about the accuracy of photogrammetric measurement may be obtained, although such information will be of a very general nature. Therefore, the characteristics of the terrain surface cover can be regarded as useful terrain descriptors.

Some parts of the terrain surface are covered by **water.** These include rivers, lakes, reservoirs, canals, etc. It may be very difficult, indeed sometimes impossible, to provide heights for these features using stereo-photogrammetric methods. Many other areas of the terrain surface are covered with **vegetation,** including forest, bush, grass, marsh, orchards, crops, as well as scattered trees, etc. These may also have their effects on the accuracy of photogrammetric measurement. Also in some areas in winter, the terrain may be covered with **snow.** Especially in high mountainous areas like some parts of North-west China, the terrain surface is covered with snow and ice the whole year round. Also in some types of terrain, the surface is covered with **rocks** or dry soils. The latter is called a **desert.** Last but not least, large pieces of terrain surface may be covered with **artificial man-made features,** e.g. roads, buildings, aircraft runways, etc. which may or may not have an influence on DTM accuracy.

3.2.2 Terrain descriptors based on the genesis of landforms

One of the terrain classification systems used in geomorphology is based on the genesis of landforms, i.e. it is related to specific processes. Generally speaking, two such forms have been distinguished as follows (Demek, 1972):-

i). endogenetic forms, which have been formed by internal forces; and
ii). exogenetic forms, which have been formed by external forces.

The **endogenetic forms** include neotectonic forms, volcanic forms and those forms resulting from deposition by hot springs. The **exogenetic**

forms include denudation forms, fluvial forms, fluvio-denudation forms, glaciofluvial forms, karst forms, suffosion forms (from Latin suffosio, meaning underwashing), glacial forms, nivation and cryogenetic forms, thermokarst forms, eolian forms, lacustrine and marine forms, forms of organic origin, and anthropogenetic forms, etc.

Of course, each form has its own special characteristics. Therefore, the accuracy of DTMs derived from sets of data (e.g. aerial photographs taken in a single run) for two adjacent areas with differing land forms could be different. This means that these descriptors can provide users with some information about the roughness of the terrain surface, and thus provide some pointers to the accuracy of DTMs. However, normally the information provided by this type of genetic classification is so general that it is difficult to incorporate it in the planning of a sampling strategy for a specific project covering a specific area, except in a very broad sense.

3.2.3 Terrain descriptors based on physiography

In a terrain classification based on physiography, an individual country, e.g. Great Britain, is divided into generalised regions according to the structure and characteristics of its landforms. Each division is kept as homogeneous as possible and each has a dominant characteristic. For example, the island of Great Britain can be divided broadly into two physiographic regions -- a Highland Zone in the north and west, and a Lowland Zone in the south and east (Stamp and Beaver, 1971).

Of course, such regions can also be sub-divided into smaller units. For example, the Highland Zone of Britain can be sub-divided into the Scottish Highlands, the Scottish Uplands, the Pennines, the Welsh Mountains, and the Southwestern Peninsula.

Such broad physiographic regions are specific to particular countries or to large regions rather than forming a classification which can be applied universally. Their purpose is primarily one of generalisation, to allow an area to be broken down into areas with consistent broad physiographic relationships, within which detailed terrain features are not specified. It is difficult to see how such broad terrain descriptors can be employed usefully in the context of DTM planning or operations.

3.2.4 Terrain descriptors based on classification of relief type and landform units

The terrain surface can also be classified at a variety of scales and with varying degree of specificity according to relief type or on the basis of land form units. The degree of detail and the nature of the

classification will vary according to the purpose involved. In such cases, the descriptors are not geographically specific in the sense of physiographic units, but are usually environmentally specific in the sense that a landform classification designed for the humid tropics will clearly be different to one designed for high altitude glacial or periglacial environments.

Similarly, the classification may involve descriptors which, by definition, are genetic in inference -- for example, a classification devised for glaciated topography involving such terms as eskers, kames, moraines, etc is, on the one hand descriptive of their surface features but, on the other hand, also implies a specific genetic origin.

Terrain descriptors based on landform units are therefore very varied in nature according to the purpose of the classification. A classification of terrain which is to be used for hydrological purposes might highlight certain aspects such as slope angles and degrees of terrain dissection. A classification intended to correlate with climatological findings might be based on height and aspect information. A terrain classification for agricultural development purposes might be based on height, dissection and rock type, so as to correlate with information on soils, drainage and vegetation.

In all these cases, the degree of accuracy and complexity will depend largely on the scale at which the classification exercise has been carried out. An exercise conducted at a scale of 1:2,500 would permit a sub-division of major terrain units into secondary terrain units and terrain facets. Conversely, a regional study at a scale of 1:100,000 would imply greater generalisation into a smaller number of terrain types.

As a basis for sampling for DTM purposes, these various types of classification would be of variable value. Small areas with a precise categorisation of landforms could well offer a satisfactory sampling frame. Conversely, a generalisation of a more descriptive nature to cover a large area would be of little value in terms of assisting a photogrammetrist in making a good decision about the sampling strategy or sampling interval to be employed over such an area.

3.2.5 Discussion of general or qualitative descriptors

Those descriptors of terrain discussed above may be so broad that they can only provide the user with some very general information about a particular landscape. Such information is often not precise enough for DTM purposes because, in this case, a mathematical model is desirable and some kind of numerical descriptors are either essential or preferred since it is numerical data that is being provided and

processed.

The terrain surface coverage gives very little information about the roughness of the terrain. For example, a bush area could be very flat or rolling. Therefore, these descriptors are, by no means, adequate for DTM purposes, though the presence of bush or forest will definitely have a negative effect on the accuracy of height measurement.

Physiographic divisions are usually too broad. For example, within the zone of the Scottish Highlands, there are still some flat areas. Of course, the roughness of these flat areas must be very different to those of the main mountainous areas. Therefore, physiographic descriptors may only be used for the general planning of DTMs over very large areas or regions, but are not adequate for an individual project.

Again with geomorphological descriptors, based either on the genesis of the landforms or the relief types and landform units, the information which they can provide will usually be too general, just like the physiographic descriptors.

Furthermore, although each of the different geomorphological types of terrain have different characteristics, the information provided by geomorphologists is so often concerned primarily with process as not to be useful to those topographic scientists who are charged with the task of DTM data acquisition. However, it must also be said that this is an area where an increased collaboration between the two groups of scientists could be fruitful in terms of the planning of DTM data acquisition and in the later modelling and reconstruction of the terrain surface.

3.3 Numerical (or quantitative) descriptors

As discussed above, the general descriptors are unsuitable for detailed DTM planning purposes although they do provide the user with some information about the terrain surface. Therefore, in this section, some numerical or quantitative descriptors such as statistical values will be discussed. The definitions of these descriptors are given at the beginning while the discussion of the usefulness of these descriptors will be given at the end under the heading of discussion.

3.3.1 Descriptors of the complexity of terrain surfaces

The complexity of a DTM surface can be specified by the concepts of roughness or irregularity. These can be characterised by different parameters such as its frequency spectrum, fractal dimension, and other geomorphometric parameters. These parameters characterise the general

shape of a terrain surface. They should, therefore, be able to serve as the basis for determining the optimum sampling strategy to be used for a specific DTM.

(1). Frequency spectrum:

A surface can be transformed from the space domain to the frequency domain by means of a Fourier Transformation. The terrain surface in its frequency domain is characterised by its **frequency spectrum.** The estimation of such a spectrum from equally spaced discrete (profile) data has been discussed by Frederiksen (1980) and his colleagues (1978). The spectrum can be approximated by the following expression:

$$S(F) = E \cdot F^{a} \qquad (3.1)$$

Where, F denotes the frequency, at which the spectrum magnitude is S(F); and E and a are constants (i.e. characteristic parameters), which are two statistics expressing the complexity of the terrain surface (or profiles) over all of the area. Thus they, like some of the numerical descriptors which will be discussed later, can also considered as rather general parameters but they do provide much more precise information about the terrain surface than those described under the heading of General or Qualitative Descriptors.

Different values could be obtained for these two characteristic parameters from different types of terrain. According to the study carried out by Frederiksen (1980), if the parameter, a, is larger than 2, the landscape is hilly with a smooth surface, and if the value of a is smaller than 2, it indicates a flat landscape with a rough surface since the surface contains relatively large variation with high frequency (or short wavelangth). Such information is, of course, very general. However, no more detailed information about the relationship between these parameters and terrain characteristics is available so far, therefore, as Frederiksen (1980) pointed out, "further experiments must be carried out to interpret the connection between the terrain variations and the spectrum".

Obviously, up till now, the use of the frequency spectrum of the terrain has not moved out of the research domain. Also a great deal of work still has to be done before the usefulness of the method is established in the context of DTMs.

(2). Fractal dimension:

Fractal dimension is another statistical descriptor which can be used to characterise the complexity of a curve or a surface. The discussion will start with the concept of **effective dimension.**

It is well-known that, in Euclidean geometry, a curve has a dimension of 1 and a surface has a dimension of 2 regardless of its complexity. However, in reality, a very irregular curve is much longer than a straight line between the same two points, and a complex surface has a much larger area than a plane over the same area. In the extreme, if a line is so irregular that it fills a plane fully, then it becomes a plane, thus having a dimension of 2. Similarly, a surface could have a dimension of 3.

In fractal geometry, the concept of which was introduced by Mandelbrot (1982), the dimension of an object is defined by necessity (meaning "practical need"), thus leading to the so-called effective dimension. The idea of effective dimension can be explained by taking the example of the shape of the Earth's surface when viewed from different distances.

i). if it is viewed from an infinite distance, the Earth appears as a point, thus having a dimension of 0;

ii). if it is viewed from a position on the Moon, it appears to be a small ball so that the terrain relief is negligible, thus having a dimension of 3;

iii). if the viewer comes nearer, for example to a distance above the Earth's surface of about 830km (the altitude of the SPOT satellite's orbit), the height information is extractable but not in detail, thus, in general terms, the observer can see a mainly smooth surface in some flat areas with a dimension of 2;

iv). if the Earth's surface is viewed on the ground, then the roughness of the surface can be seen clearly, thus the effective dimension of the surface should be greater than 2.

In fractal geometry, the effective dimension is allowed to be a fraction, which is called the **fractal dimension or fractal.** For example, the fractal dimension of a curve lies between 1 and 2, and that of a surface between 2 and 3.

The following is an example to show how the fractal dimension of a coast line can be determined (Mandelbrot, 1967). To evaluate the coast between A and B, a person may use a different measuring scale, G. For instance, if a ruler with a smallest scale (resolution) of h (i.e. G=h) is used to measure the length of this coast, then a value of L(h) (i.e. L(G)=L(h)) can be obtained. With a different value of G, he would obtain a different length L(G). The relationship between L(G) and G can be expressed as follows:

$$L(G) = M\,G^{1-D} \qquad (3.2)$$

Where, M is a constant, and D is interpreted as the fractal dimension of the coast line between A and B.

Similarly, the fractal dimension of a surface can also be measured by a series of square planes with different sizes.

From the discussion above, it can be concluded that a fractal dimension opproaching 3 indicates a very complex and probably rough surface, while a simple (near-planar) surface has a fractal dimension value which is near 2.

As in the case of the frequency spectrum approach discussed previously, the use of the fractal dimension is, as yet, far from proven to be useful in the context of defining a sampling strategy for DTM data acqusition. Much more work must be undertaken and many more measurements made before the use of fractal data becomes a viable method to be employed in DTM data acquisition.

(3). Plan and profile curvatures

It is also known that an individual relief form can be synthesized by combining form elements which are defined as relief units of homogeneous plan and profile curvatures.

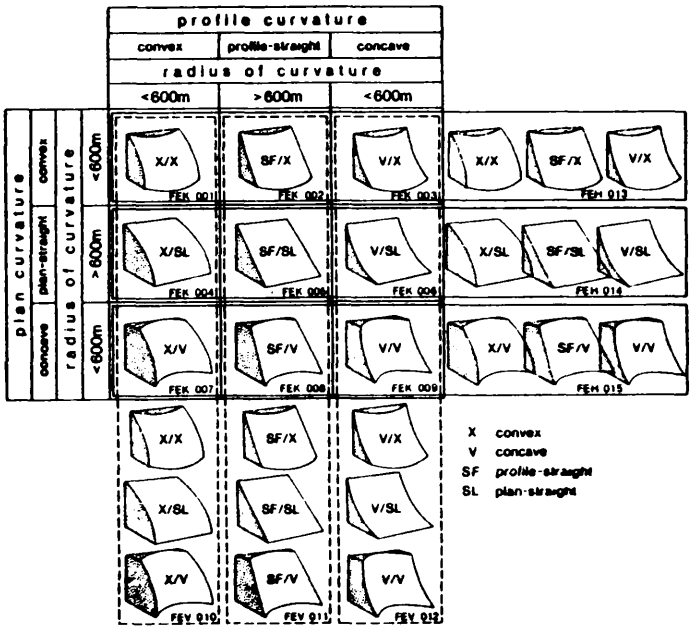


Fig.3.1 Classification of form elements by plan and profile curvature

Suppose a profile can be expressed as y = f(x), then the curvature at position x can be computed as follows:

$$C = \frac{d^2y / dx^2}{(1 + (dy/dx)^2)^{3/2}} \qquad (3.3)$$

In this respect, curvature (C) is inversely proportional to the radius of the curve (R), (i.e. C=1/R) in which case, a large value of curvature is associated with a small value of radius. Thus, intuitively, it can be seen that the larger the curvature, the rougher is the relief. Therefore, curvatures can also be used as a measure for the roughness of the terrain. This criterion has already been used for terrain analysis (Dikau, 1989). Fig.3.1 shows different type of curvatures as illustrated in Dikau's paper.

This does seems to be a potentially useful method for planning DTM sampling strategies. However, it would seem that a rather large volume of data (that of a DTM!) needs to be available to allow the curvature values to be derived -- which leads to a chicken and egg situation.

3.3.2 Descriptors for the similarity of terrain surfaces

The degree of similarity between the pairs of surface points can be described by a correlation function. This correlation function may take different forms -- auto-correlation, (auto-)covariance and variogram are all examples of these differing forms or functions. These are the descriptors of the general shape of a surface and, therefore, they can be used as the basis of weight determination in interpolation methods. For example, **variogram** has been used for this purpose in the Kriging method used in random-to-grid interpolation. They should also be able to serve as the basis for determining the optimum sampling intervals for regular gridded sampling. The computation is as follows:

Covariance: $$Cov(d) = [\ (Z_i - M)(Z_{i+d} - M)\ ] / (N-1) \qquad (3.4)$$

Auto-correlation: $$R(d) = Cov(d) / Var \qquad (3.5)$$

Where "[ ]" denotes summation; M is the mean of Zs (heights); d is the distance between two data points; N is the number of terms in the summations; and Var denotes the variance which is computed as follows:

Variance: $$Var = [\ (Z_i - M)^2\ ] / (N-1) \qquad (3.6)$$

Similarity values should have a high correlation with the complexity of the surface. The relationship should be such that the smaller the

similarity, the more complex is the surface.

(1). Covariance:

Auto-covariance, or simply covariance, is one of the concepts used to describe the similarity of a DTM surface. The formula for the computation is expressed by equation (3.4). The value of the covariance decreases with the increase of distance between pairs of data points. Some empirical models such as the **exponential model:**

$$\text{Cov}\ (d) = \text{VAR} * \text{Exp}\ (-2\ d\ /c) \qquad (3.7)$$

and the **Gaussian model:**

$$\text{Cov}\ (d) = \text{VAR} * \text{Exp}\ (-2\ d^2/c^2) \qquad (3.8)$$

have been used (Kubik and Botman, 1976), where c is the parameter indicating the correlation distances at which the value of covariance approaches to zero. Therefore, the smaller the value of c, the less similar the surface points.

The covariance can also be plotted against the distance between pairs of data points. Such a diagram is referred to as **covariogram.** The shape of the curve is determined largely by parameter "c" which has a function very similar to that of the root mean square error in the random error model.

(2). Auto-correlation:

Auto-correlation is a very powerful descriptor of a DTM surface. It is a concept similar to that of covariance. The auto-correlation function also indicates the average degree of similarity between all pairs of data points.

The auto-correlation coefficient takes a value from 0 to 1. A coefficient of 1 indicates that the elevation values of two points are identical, while a coefficient of 0 indicates that there is no similarity between the two points in terms of elevation.

(3). Variogram:

Variogram is another descriptor used to describe the similarity of a DTM surface. It has also been used for DTM accuracy estimation, and thus for the determination of optimum sampling distance. The expression for its computation is as follows:

$$2r(d) = [\ (Z(x)-Z(x+d)^2\ ]/N \qquad (3.9)$$

Where, "[ ]" denotes the summation; d is the distance between two data points; and N is the number of summation terms. Then r(d) is the value of the semi-variogram between data points with a distance of d apart. In the elevation case, the following model is often suitable:

$$r(d) = A * d^b \qquad (3.10)$$

Where, A and b are constants; and d is still the distance between two data points. This model is known as the De Wijsian model (see David, 1977).

It can also be concluded that the smaller the value of the variogram, the greater the similarity between two sampled points in terms of elevation.

3.3.3 Discussion of numerical descriptors

These numerical descriptors are essentially statistical descriptors. They are computed from a sample of terrain points from the project area. Usually, some profiles are used as the sample and then a numerical value is computed from these profiles. However, there are some problems connected with this approach. One of these is that the numerical value computed from the selected profiles could be different to that derived from the whole surface. If one tries to compute it for the whole surface, then a sample from the whole surface is necessary. In this case, the original purpose of having a terrain descriptor for project planning disappears. Also the numerical values are very sensitive to the sampling interval used, the length of profile used and even the location and direction of the profiles used in the computation. Therefore, it is very possible that the obtained numerical descriptor fails to convey the correct message to a DTM planner.

In conclusion, it is often very difficult to have a robust estimate of these numerical descriptors for a given area, thus it is difficult to make use of them in defining a sampling strategy. Thus in practical terms, these descriptors have been used experimentally in DTM research and practice but not so far successfully in operational or routine production terms.

3.4 A recommendation for slope together with wavelength to be used as the main terrain descriptor for DTM purposes

It has been found that these various numerical/statistical descriptors discussed in the previous section are not too suitable for the purposes of DTM planning and practice. Therefore, the aim of this section is to

recommend a more practical approach.

3.4.1 Roughness vector: Slope and wavelength

Roughness is the concept originally used in geomorphometry to describe the complexity of a topographic surface. The parameters used for such purposes have been reviewed by Mark (1975). It has been found that roughness cannot be completely defined by any single parameter, but must be represented by a **roughness vector** or set of parameters.

In this set of parameters, **relief** is used to describe the vertical dimension (or amplitude of the topography), while the terms **grain** and **texture** (the longest and shortest significant **wavelengths**) are used to describe the horizontal variation. The parameters for these two dimensions are connected by **slope.** Thus, relief, wavelength and slope are the roughness parameters. The relationship between them can be illustrated in Fig.3.2(a). It can be seen that the slope angle at a point on the "wave" varies from position to position. The following mathematical equation may be used as an approximate expression of their relationships:-

$$\tan A = H / (W/2) = 2 H / W \qquad (3.11)$$

Where, A denotes the average value of the slope angle; H is the local relief value; and W is the so-called wavelength. It is clear that, if any two of them are known, then the third can be computed from Equ.(3.11).

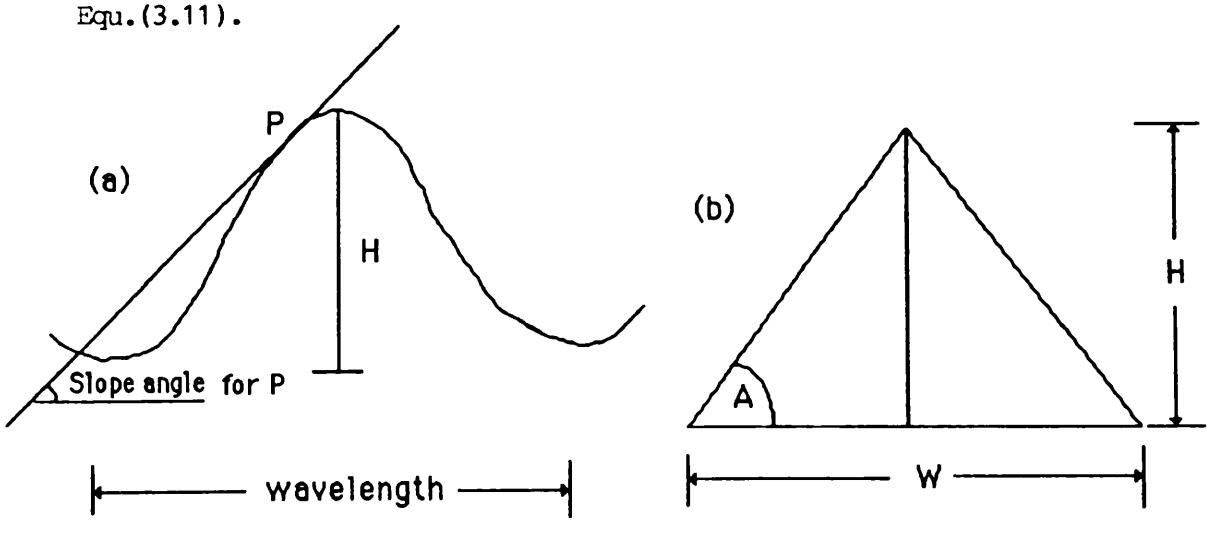


Fig.3.2 Relationship between slope, wavelength and relief (a). Their full relationship; and (b). simplified diagram

For many reasons which will be discussed in the next sections, **slope and wavelength** together are recommended as the main terrain descriptors for DTM purposes.

In practical terms, the relief parameter for a given area will often be known. Therefore, the value of the corresponding wavelength will also be known if the slope value is known. Therefore, the adequacy of this descriptor depends above all on the availability of slope values in terms of terrain description.

3.4.2 The adequacy of slope as a terrain descriptor from the geomorphometric point of view

Evans (1981) states that "a useful description of the landform at any point is given by altitude and the surface derivatives, i.e. slope and convexity (curvature)." "**Slope** is defined by a plane tangent to the surface at the given point and is completely specified by the two components: gradient (vertical component) and aspect (plan component)." "**Gradient** is essentially the first vertical derivative of the altitude surface while **aspect** is the first horizontal derivative." Further land surface properties are specified by **convexity** (incluing positive and negative convexity - **concavity**). These are the change rates in the gradient at a point (in profile) and the aspect (in plan tangential to the contour passing through this point). In other words, they are the second derivatives.

These five attributes (altitude, gradient, aspect, profile convexity and plan convexity) are the main elements which are descriptive of surface points. Among them, slope, comprising both gradient and aspect components, is the fundamental attribute.

The gradient component should be measured in the steepest direction. However, when taking the gradient of a profile or in a specific direction, it is actually the vector of gradient and aspect that is obtained and used. Therefore, the term slope or slope angle is used in this thesis to refer to the gradient in any specific direction.

3.4.3 The adequacy of slope as a terrain descriptor from other different points of view

The usage of slope as the main terrain descriptor for DTM purposes can be justified for the following reasons:

i). Slope is a very powerful terrain descriptor. As quoted by Evans (1972), Strahler (1956) pointed out that "slope is perhaps the most important aspect of surface form, since surfaces may be formed completely from slope angles..."

ii). Slope is the first derivative of altitude on the terrain surface. It shows the rate of the change of height of the terrain surface with distance.

iii). Traditionally, slope has been recognised to be a very important descriptor and is well utilized in practice. For example, the map specifications for contours are given all over the world in terms of slope angle.

iv). More importantly, in DTM practice, Ley (1986) found that "a **positive correlation** exists between the vertical error of a regional DTM and the mean slope of that region as measured from the DTM. The correlation coefficient was very high with a low standard error". From this discovery, Ley (1986) then concluded that it is possible to predict the vertical accuracy of DTMs purely by analysing the mean slope of the model".

3.4.4 Practical considerations for slope

In the previous section, it was suggested that slope could be used as the main terrain descriptor for DTM purposes. In this section, some of the problems and/or difficulties of using slope will be discussed.

One of the problems is the **availability** of slope data. In the present context, the purpose of classifying the terrain on the basis of slope values is to design a sampling strategy for this area. Therefore, the slope values should be available or estimated before sampling takes place.

Actually, slope can be determined directly from the height and plan measurements carried out in a stereo-model or derived from contour maps. The method designed by Wentworth (1930) is still widely used for estimating the average slope of an area from a contour map. If there is no contour map available for the project area, then the slope may be estimated from aerial photographs. Some of the methods which are available for the measurement of slope from aerial photos have been compared by Turner (1977).

Another problem may be the **variability** of slope values. Slope may vary from place to place so the slope estimate which is representative for one area may be not suitable for another. In this case, the average value may be used in suitable situations as suggested by Ley (1986). If the slope varies too greatly in an area, then the area may be divided into small pieces and different sampling strategies may be applied to each of these pieces. However, the discussion of the design of this sampling strategy will take place later (i.e. outside of this chapter) after the discussion of the accuracy of DTMs from different data sources.

Although there may be some difficulties in measuring or estimating

slope, just like those encountered in providing estimates of other numerical descriptors, there are still some advantages of slope over the others:

i). One of these is that people have a lot of experience in dealing with slopes when making topographic maps. Thus a lot of prior knowledge is available that can be put to use in defining or deciding upon a sampling strategy.

ii). The second is that, there is evidence of a high correlation between slope and relief (e.g. see Evans, 1972). Therefore, relief information may be used as a rough guide for the value of slope. For example, in deciding the contour intervals to be used in small scale topographic mapping, Imhof (1965) takes 45^{o} as the average slope value used in for high mountainous areas of rugged relief, 26^{o} for lower mountainous areas with a less rugged relief and 9^{o} for relatively flatter areas.

iii). Due to the high correlation between slope and relief, the concept appears very intuitive to every one, though obviously there are exceptions such as high plateaux with large elevation or altitude values but maybe a relatively gentle terrain in terms of slope.

3.5 Concluding remarks

In this chapter, a brief discussion of terrain descriptors has been made. These descriptors include general descriptors and numerical descriptors. It has also been pointed out that the general or qualitative descriptors can only provide users with such a general knowledge about the roughness of the terrain surface that they can only be used in the planning stage for modelling large areas. By contrast, numerical or quantitative descriptors are capable of providing more precise information about the roughness of the terrain. Thus they can be used at the design stage for specific projects within a relatively small area. However, there are some problems associated with those numerical descriptors which have already been used in DTM practice. Due to these problems, it is difficult to make practical use of these descriptors.

From both the theoretical analysis and the results from DTM practice, slope together with wavelength appears to be a very promising terrain descriptor, and therefore it will be the main descriptor used in this project.

Using numerical parameters to classify terrain is referred to as a

parametric classification. In this system, terrain is classified on the basis of the user's demand for a specific purpose -- in this particular case, the acquisition of DTM data. In the case of terrain modelling, the main purpose is to decide upon and then design the sampling strategy, the basis for which will be discussed in the next chapter.

Chapter 4

Sampling Strategy for Data Acquisition Using Photogrammetric Methods

Chapter Four

Sampling Strategy for Data Acquisition Using Photogrammetric Methods

After the classification of terrain types, the next step is data acquisition. Since the formation of a digital terrain model is a numerical and mathematical process, a suitable set of digital coordinate data of the terrain surface is required before any modelling can be carried out. The acquisition of this data is the primary (and probably the single most important) stage in the terrain modelling process. This is the topic of this chapter.

4.1 Introduction and background

The source data, comprising positional coordinates and elevation values, can be measured directly on the terrain surface by field survey methods or it can be obtained indirectly from other sources such as aerial photographs, remotely sensed imagery, and/or existing contour maps. The method used will depend partly on the availability of these different materials but also on the scale or the required sampling interval and the accuracy requirements of the DTM. Furthermore, the intended use and the type of information which will be extracted from the model will help to decide which source and which particular method will be used to acquire the data for the DTM.

4.1.1 Data sources for digital terrain modelling

The actual terrain surface of the Earth occupies a vast area -- about 149.5 million square kilometres -- although it is only a small fraction (29.2%) of the total area of the Earth's surface. The approach of acquiring data directly from measurements on the terrain surface is referred to as ground survey or field survey.

The data acquired by **ground survey** is (or should be) of a very high quality in terms of accuracy. It is suitable for large-scale terrain modelling such as projects in civil and mining engineering covering a relatively limited areal extent. However, it is not an efficient or practical approach for measuring or modelling large areas of the Earth's surface. Therefore, consideration of digital terrain modelling based on data from this particular source will not be undertaken in this thesis.

Space photographs are another source which could be used for terrain modelling. Some photographic systems such as the Skylab S-190A and S-190B, the Spacelab Metric Camera and the Large Format Camera have been used experimentally to collect terrain model data. However, the height

accuracy of data points measured from the photographs taken by these systems is so low that they can only be used for reconnaissance purposes. **Remotely sensed imagery,** taken from space can also be considered as a data source for terrain modelling. However, although many satellite scanner systems such as the MSS and TM sensors used in Landsat series and the stereo-scanner used in SPOT have been experimented with, the quality of height information which can be obtained from such systems is very poor in terms of both relative and absolute elevation accuracy. Thus, it is impractical to use such space imageries as the data source for terrain modelling except for some kind of reconnaissance purpose. Therefore, least consideration has been given in this thesis to using this second type of space image as a data source. However, a test of DTM accuracy has been carried out using space photography and is reported in Chapter 10.

Existing contour maps are another source for terrain modelling. The acquisition of data from existing maps is referred to as cartographic digitisation. The problems associated with this particular data source will be discussed later in Chapter 6 and a practical test of accuracy of a DTM derived from this type of data will be reported in Chapter 13.

Aerial Photography is a vital and very useful data source for many different disciplines. It has been widely used in topographic mapping, geology, agriculture, forestry, ecology, archaeology, resource management, urban planning, etc. It is also the main source for the production of accurate DTMs over large areas and so its characteristics will be considered at length in this thesis.

4.1.2 Photogrammetric methods of DTM data acquisition

The method used to acquire (i.e. measure) data from aerial photographs is referred to as photogrammetry. Photogrammetric data can be acquired for DTM production in any of several different ways. It is possible to use a photogrammetric operator to measure a set of data points on a stereo-model following some kind of procedure designed specifically for the purpose, or it can be collected as a by-product of some other photogrammetric operations such as orthophoto production or contouring. It is also possible to obtain photogrammetric data by automated or semi-automated measuring methods, e.g. using correlation techniques.

The specific procedure designed for the purpose of DTM data acquisition is referred to as the **photogrammetric sampling method** in this thesis and forms the principal topic of this chapter.

Indeed, sampling is a vital matter in digital terrain modelling since more and more evidence is becoming available which shows that the loss of fidelity in terrain topography resulting from the failure to

acquire sufficient or appropriate information at the data aquisition stage cannot be regained by interpolation at the reconstruction or modelling stage (see Makarovic, 1972; Kubik and Botman, 1976; Ackermann, 1979). Therefore, determining the location of points to be measured and/or the optimum sampling interval between points is the most important topic in sampling theory. This will be discussed later in more detail in Chapter 14 which is concerned with the optimization of photogrammetric sampling for different types of DTM projects after discussion of the accuracy of digital terrain models.

Various photogrammetric sampling methods have already been in use. These include profiling, regular grid sampling, selective sampling (Makarovic, 1984), progressive sampling (Makarovic, 1973) and composite sampling (Makarovic, 1977). However, the theoretical background of these different methods, the effectiveness of each method and the overall philosophy of the photogrammetric sampling methods themselves have rarely been discussed. However, such a discussion is of great importance if the most suitable sampling method for a specific project is to be selected. Also the discussion might lead to the development of more comprehensive and effective methods.

The concept of sampling used above refers to the sampling before measurements, thus it can be considered as **pre-measurement sampling.** However, in some cases, the data points may have already been measured with a great density, as in the case of data sets produced by automatic correlation, but not all points may be necessary for a specific project. Therefore, some methods of filtering need to be applied to such data sets. This type of filtering can be called **post-measurement filtering** and a brief discussion on this topic will be given at the end of this chapter.

4.2 Theoretical background for photogrammetric sampling

From the theoretical point of view, a **point** on terrain surface is zero-dimensional, thus without size, while a terrain **surface** comprises an infinite number of such points. If full information about the geometry of a terrain surface is required, it is necessary to measure an infinite number of points. That is to say, from both the theoretical and practical points of view, it is impossible to obtain full information about the geometry of a terrain surface.

However, in photogrammetric practice, the terrain surface is represented quite adequately by the surface of the reconstructed stereo-model of the area being modelled. It follows that the height of a point measured on the stereo-model surface (which is actually formed from two tiny fuzzy circles on aerial photographs) represents the elevation value

(height) over a certain size of area. Any lack of accuracy in this value is mainly due to the **limitations of the photographic imaging system** (including the resolution of the aerial camera-lens/emulsion combination) and to the mechanical accuracy of the photogrammetric instrument.

In practice, a stereo-model surface is represented by a finite number of measured points. That is to say, a stereo-model (and thus a piece of the terrain surface) can be represented quite adequately by measuring only a finite number of points on it. This fact makes photogrammetric sampling meaningful in a practical sense.

In most cases, full or complete information about a stereo-model surface is not required for a specific DTM project, and so the photogrammetric measurements required for a DTM project can be carried out with a view to obtaining a set of digital data which can represent the corresponding terrain surface to the required density and/or to the required degree of accuracy and fidelity.

4.3 Sampling theorem and its possible application to DTM data sampling

The fundamental sampling theorem, used in mathematics, statistics, engineering and other related disciplines is well-known and can be stated as follows:

> "If a function is sampled at an interval of Dx, then the variations at frequencies higher than 1/(2 Dx) cannot be reconstructed from the sampled data."

As Tobler (1969) explained, the theorem states that "if a function has no spectral components of frequency higher than W, then the value of the function is completely determined by a knowledge of its values at points spaced 1/(2W) apart."

In the case of terrain modelling, if a terrain profile is long enough to be representative of the local terrain, it can then be represented by a sum of its sine and cosine waves. If it is assumed that the number of the terms in this sum is finite, there is, therefore, a maximum frequency value, say F, for this set of sinusoidals. According to this theorem, the terrain profile can be completely constructed if the sampling interval along the profile is smaller than 1/(2F). Therefore, extending this idea to surfaces, the sampling theorem could also be used to determine the sampling interval between profiles in order to obtain adequate information about a terrain surface which is represented by a stereo-model surface.

Furthermore, according to this theorem, if a terrain profile is sampled at an interval of Dx, then the information about the variations of the profile with a wave length of 2 times Dx or more is totally lost. Therefore, as Peucker (1972) has pointed out, "a given regular grid of sampling points can depict only those variations of the data with wave lengths of twice the sampling interval or more."

This theorem has been used for the determination of the sampling interval by Frederiksen (1980) and his colleagues (Frederiksen et al, 1986). The computation takes place in the frequency domain.

4.4 Photogrammetric sampling from different points of view

The surface of a stereo-model can be considered to be an assemblage of points. These points can be viewed in various ways from the differing view points inherent in subjects such as statistics, geometry, topographic science, etc. Therefore, different sampling methods can be designed and evaluated according to each of these different view points. The following paragraphs will discuss these view points and methods.

4.4.1 Statistics-based sampling

From the view point of statistics, a stereo-model surface is a **population** (also called a sample space) and the sampling can be carried out either randomly or systematically. The population can then be studied by the sampled data.

Random sampling is sampling in which any sampled point is selected by a chance mechanism with known chances of selection. The chance of selection may be different from point to point. If the chance is equal for all sampled points, then it is referred to as **simple random sampling.** In **systematic sampling,** the sampled points are selected in a specially designed way, each with a chance of 100% probability of being selected.

The whole population can be sub-divided into several sub-populations, each called a **stratum,** and the sampled points can be taken from each of these strata. Any sampling carried out in this way is referred to as **stratified sampling.** The sampling can also be taken in another way in which the whole population is divided first into several groups (referred to as clusters in statistics), and then one of them is randomly selected for the estimation of the population. This method is referred to as **cluster sampling.**

4.4.2 Geometry-based sampling

From the geometrical point of view, a stereo-model surface can be represented by different geometric patterns. This pattern could be either irregular or regular in nature. The regular pattern can, in turn, be sub-divided into 1-D regular and 2-D regular patterns.

The data points obtained from the sampling with an **irregular pattern** can be represented by irregularly-shaped triangles or polygons. If the sampling is conducted with a **regular pattern** which is only regular in one dimension, then the corresponding method is referred to as **profiling (or contouring).** A two-dimensionally regular pattern could be a **square or regular grid,** or a series of contiguous equilateral triangles, hexagons, or other regularly shaped geometric figures.

4.4.3 Feature-based sampling

In this thesis, "feature-based sampling" refers to the sampling as seen from the point of view of topographic science. From this view point, a stereo-model surface consists of a finite number of surface points, and the information content of these surface points may vary with their positions on the stereo-model surface (representing the corresponding **terrain** surface). In surface theory, surface points are classified into two groups, one of which comprises feature-specific points (and lines) while the other comprises random points.

Feature-specific points are those points on the surface which represent surface features with a much higher or more significant information content than the average surface point. Examples of these are peaks, pits and passes. These points not only contain their own coordinate values, but are also be able to give some information about their surroundings. **Peaks** are the summits of mountains and hills, so they have a set of points of lower height around them. By contrast, **pits** are the bottoms of valleys (or holes), so they have a set of greater height values around them.

The lines connecting certain types of feature-specific points are referred to as **feature-specific lines** in this thesis. Examples of these are ridge lines and course lines (e.g. rivers, valleys, ravines, etc.). **Ridge lines** are the lines connecting pairs of points such that the points on them are local maxima; and **course lines** are, on the other hand, linking pairs or strings of points so that the points defined by them are local minima.

The crossing points of these two types of lines are referred to as **Passes.** They are, therefore, the points which, at the same time, can have a maximum elevation value in one direction and a minimum value in

the other direction.

As discussed in Chapter 3, it has been found in geomorphometry, that a terrain surface may be characterised almost completely by its slope angles. From this point of view, it can be thought that the importance of feature-specific points present on the terrain surface arises from the fact that, at these points, terrain slope changes not only its direction but also its sign or magnitude. For example, at peaks, it changes from positive to negative and at pits, it changes from negative to positive. There are also two other types of points at which the terrain slope changes its vertical angle but not its sign. They are **convex points** and **concave points.** If a slope is viewed as an up-down transition, then the slope change is from gentle to steep at the former and in an opposite direction at the latter.

The convex and concave points are also, invariably feature-specific points, connected to become linear features. If there is a special case where the slope change is very sudden, then these linear features are referred to as **break lines.**

All these points -- feature-specific points, points along feature-specific lines, points along break lines, etc. -- are very important points (VIPs) in representing the terrain surface.

As has been discussed before, in the statistics-based sampling approach, every surface point is considered as a random point. However, in the topographic science context, a **random point** (RP) refers to any point randomly located on the surface. Unlike the VIPs, a random point which has been measured can only give information about its own height value but nothing more. That is to say, if a surface may be represented by N points, knowing one of them only gives us 1/N more information about the geometry of the surface. From the point of view of topographic science, random points are of only minor significance and little effort is made to collect data on this basis.

The discussions above suggest that there is very little loss of information about the terrain surface features if any points other than VIPs are not selected. In other words, if only these VIPs are selected, the main features of the terrain surface can still be obtained. This results in the method of **selective sampling.** In this context, any other sampling method can be classified into the group of **non-selective sampling.**

4.5 The requirements for photogrammetric sampling methods

In the preceeding section, photogrammetric sampling has been discussed

from several different points of views. This section will discuss the requirements for a photogrammetric sampling method.

The whole DTM production system can be considered as an engineering system. The following criteria are the main standards for system judgement: performance, cost, reliability and time (Chestnut, 1965). Specifically for a sampling method, the requirements are described as follows:

i). In order that the whole DTM production system can have a good **performance**, the sampling method should enable data to be obtained with sufficient **fidelity** to recreate the original surface while fitting into a comprehensive **data structure.**

ii). As far as the **time** factor is concerned, a sampling method is required which will enable the operator to measure a stereo-model rapidly and have the sampled data processed very efficiently.

iii). **Cost** is another concern for a DTM production system and must include the cost of purchasing and operating the hardware, the cost of accommodation, personnel costs including salaries and insurance, etc. The cost factor should also include items such as the CPU time of data processing. In addition, a sampling method should enable the operator to measure data points in as straightforward a way as possible in order not to take excessive time for measurement nor to incur the cost of additional hardware and/or software.

iv). The discussion above is about the general requirements for a photogrammetric sampling method. In practice, a further matter of high concern is with the **structure** of the sampled data. This is related to the characteristics of the modelling programs. For example, if a surface modelling program accepts only square grid data sets, then as far as possible, the (square) grid sampling method should be considered since the use of any other pattern will lead to the need for substantial pre-processing, including the transformation of and interpolation from the measured data before it can be employed in the modelling program.

4.6 Existing photogrammetric sampling methods

The general requirements for a sampling method have been discussed in previous sections. This section will discuss the merits and demerits of existing sampling methods so that other potential sampling methods can be searched for.

(1). Selective sampling

Using this method, the data on the stereo-model is obtained in a similar way to the procedure used in optical and electronic tacheometry in land surveying. That is to say, all the VIPs are selected, thereby ensuring that the data is reasonably comprehensive in its coverage. This method has the distinct advantage that fewer points can represent the surface with high fidelity. But it is not an efficient way of selecting data points because it requires substantial interpretation of stereo-model by a trained human observer. In practice, no automated procedure can be implemented on the basis of this strategy so that it is not popular in certain mapping organisations (e.g. military survey organisations) where speed of data acquisition is of prime importance.

(2). Regular grid sampling

As the name implies, regular grid sampling ensures that the data points are obtained in the form of regular grid. This can be achieved by keeping one dimension (e.g. the X direction) constant and changing position along the line of the second dimension (Y in this case) in a series of equidistant steps while measuring the height (Z) values at each grid node. In this method, a microcomputer can be implemented to control the movement in the both X and the Y directions instead of it being implemented manually. That is to say, it is convenient for the automatic or semi-automatic positioning of the data points which have to be measured. Another advantage of using this method is that the height data can be stored in a regular matrix, which simplifies the processing of data. For these reasons, this is a popular method in many surveying and mapping organisations.

But in terms of sampling, a heavy redundancy of data is required in order to ensure that all slope discontinuities are detected or the changes in the topographic surface texture are represented in an adequate manner.

(3). Progressive sampling

In order to solve the problem of data redundancy in regular grid sampling, Makarovic (1973) designed a modified method, which he called progressive sampling. In this procedure, the sampling is carried out in a grid pattern whose interval changes progressively from coarse to fine over a small area.

The procedure is: First, a set of grid points are measured at a low density, then the elevation values at these data points are analysed by

an on-line microcomputer. In turn, the computer generates the location of new points which are required to be sampled in next run. The procedure may be repeated until some prior criteria are satisfied.

For such criteria, Makarovic (1973) proposed initially to use the second differences of elevation values computed along both rows and columns of the measured (sampled) coarse grid. Several additional or alternative criteria have also been proposed by Makarovic later (1975), such as the so-called random-variation criterion, parabolic criterion, distance criterion and contour criterion, etc. Of course, these other criteria may also be used as the basis of the sampling strategy for a particular terrain type. However, this is by no means the principal topic of this section.

Progressive sampling can solve part of the redundancy problem that is inherent in the regular grid sampling method, but still there are shortcomings as Makarovic (1979) has noted:

i). the sampled data points exhibit a high degree of redundancy in the proximity of abrupt changes in terrain surface;

ii). some pertinent features might be lost in the first run with its wide (coarse) spacing. These cannot be recovered by the following sampling runs; and

iii). the tracking path is rather long, which incurs a decrease of the time efficiency.

To overcome these drawbacks of this method, Makarovic (1977) proposed the use of another method which he called composite sampling.

(4). Composite sampling

Composite sampling is the combination of selective and regular grid sampling or a combination of selective and progressive sampling. Abrupt changes -- specific terrain features in the terrain surface such as ridges, break lines, etc. -- are sampled selectively, and these values and feature-specific points -- peaks, passes and hollows -- are added to the regular grid-sampled data.

The use of this method may solve many of the problems encountered with regular grid sampling and/or progressive sampling.

(5). Profiling

This method is similar to regular grid sampling. The only difference between them is that the measured elevation data yielded by regular

grid sampling is regular in both directions of the grid cells, whereas in profiling, it is regular in only one-dimension along the direction of profiling. Usually, the data points are measured in a dynamic mode rather than in the static mode which is normally associated with regular grid sampling. Thus the actual measurement process should be very efficient in terms of speed. But the accuracy of the measured data will be lower than that of the data obtained from regular grid sampling where the height measurement is carried out in a stationary position. A problem with the profiling method is the large data redundancy which is inherent in the method if small but important terrain features are to be picked up.

Frequently,this method is associated with orthophotograph production rather than being carried out primarily for the production of DTM data. In this sense, the DTM data can be regarded as a by-product rather than a main product of the profiling process. Users must therefore view the shortcomings of this method (e.g. in terms of height accuracy) and the potential usefulness of the measured data, bearing this in mind.

4.7 Attributes of photogrammetric sampling

In the context of digital terrain modelling, sampling is the process of selecting those points which have to be measured in certain positions. This operation can be characterised by two parameters: the distribution (location and pattern) and the density of the points whose elevations have to be measured. These two parameters are referred to as the attributes of sampling in this thesis. This section is to discuss the problems associated with these attributes of sampling.

4.7.1 Distribution of sampled data

The distribution of sampled data is usually specified by the terms "location" and "pattern". The **location** will of course be defined in terms of two positional coordinates - longitude and latitude in the geographical coordinate system, or easting and northing in grid coordinates. Regarding **pattern**, a variety of these, such as a regular or rectangular grid, are available for selection. Some discussion about these patterns has been already given in the section on "geometry-based sampling". A further discussion starting from the point of view of the data itself together with existing sampling methods will be given in this section.

These patterns can be classified in different ways, as one likes. The Figure 4.1 shows one way of carrying out such a classification.

Regular two-dimensional data is produced by means of regular grid

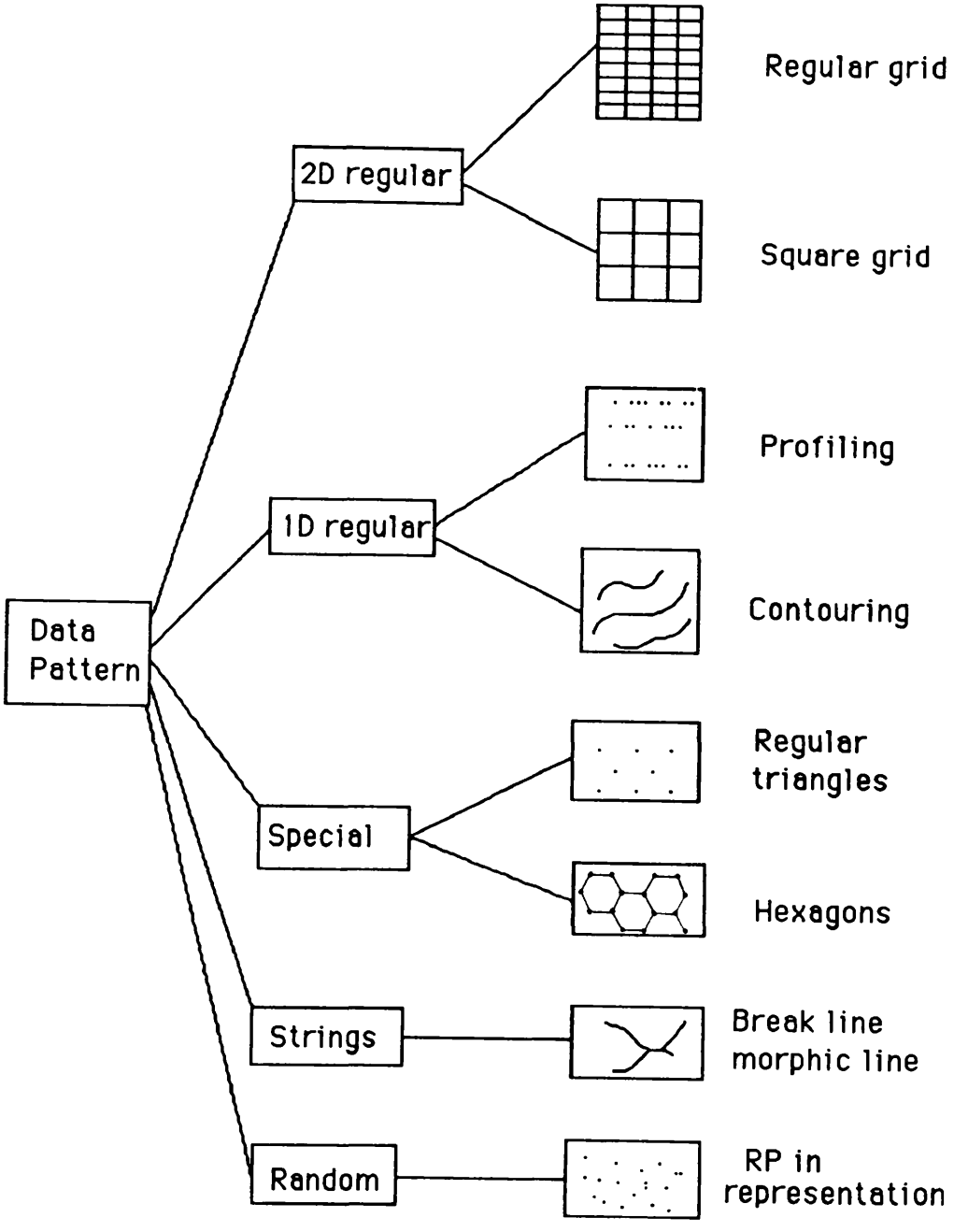


Fig.4.1 Patterns of photogrammetrically sampled data

sampling or progressive sampling. The resulting pattern could be a rectangular grid, a square grid, or a hierarchical (or progressive) structure of these two. The square grid is most commonly used. The hierarchical structured data, sampled by means of progressive sampling can be decomposed into a normal square grid structure.

Data which is regular in one dimension is produced by sampling with one dimension fixed (in X, Y or Z). When this operation is carried out for either of the first two values being fixed, it is referred to as **profiling**; if the Z value is fixed, then it is called **contouring.**

There are also other special regular patterns, e.g. equilateral triangles and hexagons, etc. However, it seems to the author that these structures are not in any way as widely used in DTM practice as profiled data or the use of a regular (or square) grid.

As has been discussed before, data patterns can be divided into two categories, i.e. regular and irregular patterns. Regular patterns have been discussed above. Regarding irregular patterns, they may be generally classified into two groups, i.e. random data and string data. By **random data,** it is meant here that the measured points are located randomly, i.e. not in any specific way.

Another irregular data pattern is that produced in the form of **strings.** Although string data is not located in a regular pattern, yet it does follow certain features (e.g. break lines). The data sets which are sampled along rivers, break lines, as well as morphological lines, etc. all belong to this pattern. This is actually not an independent pattern, but rather a supplemental one which is feature-specific. For example, the pattern of the data resulting from composite sampling is usually a combination of string data with regular (or rectangular) gridded data.

4.7.2 Density of sampled data

Density is another attribute of sampled data. It can be specified by a few measures, such as the distance between two points, the number of points per unit area, the cut-off frequency, etc.

The distance between two sampled points is usually referred to as the **sampling interval** (or sampling distance or sample spacing). If the sampling interval varies with position, then an average (or representative) value should be used. This measure is specified by a number with an unit, e.g. 20 metres. Another measure which could be used in DTM practice is the number of points per unit area, e.g. 500 points per square kilometre.

If the sampling interval is transformed from the space domain to the frequency domain, then the **cut-frequency** (the maximum frequency which the sampled data represents) can be obtained. From other points of view, the required **maximum frequency** can also be used as a measure of data density because the sampling interval can also be obtained from it (the value of maximum frequency).

4.8 The concept of post-measurement filtering

As mentioned in the introductory section of this chapter, post-measurement filtering is also necessary in some cases. For example, if the data sets are derived from digital image correlation techniques, e.g. obtained automatically from correlator-equipped analytical plotters, then the resulting data points could be very dense -- 500,000 to 700,000 points per stereo-model in the case of the GPM-2 (Petrie, 1990). Such dense data sets are not always an advantage but instead, in some cases, constitute a definite drawback. Therefore, an optimum density of data points should be used in the surface modelling process -- a matter which will be discussed in Chapter 14 of this thesis. The procedure used for the reduction in data density from the original data set to the optimum one can be referred to as **post-measurement filtering.**

Post-measurement filtering is, actually, the selection of data points from the very dense set of measured points which define the terrain surface. For such a filtering to be successful, some preliminary analysis regarding the information content of the data points needs to be carried out to determine which points should be selected. In order to do this, certain selection criteria, for example based on the second derivative values and/or curvatures, must be applied to the data set. For instance, any point with a profile curvature greater that 1/600 might be considered as being significant and included in the reduced (sampled) data set. In this way, the density of data points will be reduced but the information loss from the data sets is kept to a minimum.

If the original data is irregularly distributed, then local surfaces may need to be reconstructed for the computations of curvatures at data points.

The above discussions are included at this stage only for the introduction of the concept of post-measurement filtering. In the published literature, a different terminology may be used including such terms as data post-processing, redundancy removal, etc. However, essentially, these are covering the same basic concepts and procedures.

4.9 Concluding remarks

In this chapter, the theoretical background for photogrammetric sampling has been given; sampling strategies using photogrammetric methods have been examined; the requirements for sampling methods have been sketched; existing photogrammetric sampling methods have been reviewed; the attributes of photogrammetric sampling have also been discussed; and at the end, the concept of post-measurement filtering has been introduced.

Data acquisition using photogrammetric methods consists, as we know, of two main operations, i.e. sampling and measurement. A discussion of photogrammetric sampling has already been presented in this chapter, so following on in the next chapter should be a discussion on photogrammetric measurements carried out for DTM purposes, including the especially important matter of the accuracy of photogrammetrically measured data.

Chapter 5

The Accuracy of Photogrammetrically Measured Data

Chapter Five

The Accuracy of Photogrammetrically Measured Data (PMD)

In DTM data acquisition using photogrammetric methods, two main operations are involved. One of them is the sampling and the other is the actual measurement. The former can be characterised by two parameters, i.e. the distribution and density of the data while the latter can be described by its accuracy. Photogrammetric sampling has been discussed in general terms in the previous chapter, so it is the task of this chapter to discuss photogrammetric measurement and the accuracy of the resulting data.

5.1 Introduction

The errors which occur in photogrammetrically measured data (PMD) are the consequence of the errors involved in all the operations and materials used in the process. Broadly speaking, they comprise the errors inherent and present in the stereo-model itself and those involved in the actual photogrammetric measurements of position and height. Thus, it can be said that the overall accuracy of PMD is a function of the fidelity of the stereo-model and the accuracy of photogrammetric measurement.

The **fidelity of the stereo-model** itself is determined by the quality and metric characteristics of the actual aerial photographs from which the model is formed; the accuracy of the control points provided by ground survey or aerial triangulation; the skill and care taken by the operator in the orientation procedure; as well as the accuracy of the photogrammetric instrument itself.

To large extent, the **accuracy of measurement** is also determined by the skill and experience of the operator, the state of the adjustment of the photogrammetric instrument and the morphological characteristics of the terrain, especially its slope and roughness.

Many tests of photogrammetric accuracy have already been carried out by several international organisations, especially ISPRS and OEEPE, and some of these results will be reviewed in this chapter. However, first of all, a short theoretical discussion must also be provided for the convenience of later presentation and discussion.

5.2 The accuracy of photogrammetric instruments

From the discussion in the previous section, it can be seen that the

accuracy of the photogrammetric instrument which is used for data acquisition affects both the fidelity of the stereo-model and the accuracy of photogrammetric measurement. Therefore, it seems pertinent to allocate a complete section for the discussion of this matter.

The **accuracy of photogrammetric instruments** is usually specified for a particular type of instrument but not for an individual example of that type. For an individual stereo-plotting machine, if it is known that it belongs to a particular type or class of instrument, then its accuracy will be known in general terms. However, the individual instrument's accuracy will not be known in exact terms and this can only be established by carrying out some form of calibration procedure, e.g. tests conducted using high precision grid plates.

A great variety of photogrammetric instruments is currently available with very substantial differences between them in terms of concept, design, construction, and expense. The classification of these instruments can be based on different criteria. A very basic classification is into analogue, analytical and digital instruments. In the first two groups, hard copy photographs are used to form the stereo-model, while digitized photographs are used in the third.

With **analogue instruments,** a physical reprojection of the ray directions which existed at exposure is implemented, e.g. using optical rays or mechanical rulers or space rods.

Analogue stereo-plotters can be sub-divided into a few groups according to some criteria. Two such systems are in common usage. The first is based on the actual type of physical projection system used -- optical, mechanical, optical-mechanical, etc. This is supplemented by a statement of its order, i.e. first order, second order and third order stereo-plotters, a classification based on both the accuracy and the range of functions which can be carried out by the measurement. An alternative to this system, based on the same criteria, led to their classification into universal, precision, topographic and approximate instruments.

In **analytical and digital instruments,** the reprojection and model formation is purely mathematical and invariably involves the use of a high-speed computer. Analytical instruments may be sub-divided into comparators, image space plotters and analytical plotters with either image or object coordinates as primary input to the computational process. Till now, only one or two digital instruments have appeared, so a classification of these instruments is inappropriate at present.

Associated with all such classifications is a further classification based on the degree of **automation.** Using this criterion, the photogram-

metric instruments can be classified into three categories: i.e. fully automated, partly automated, and manual stereo-plotting instruments.

The **accuracy** of photogrammetric instruments is normally tested using precise grid plates which have been calibrated by a standards laboratory. Such a test result is an overall result, affected by many factors such as the projector system, measuring system, etc. Two types of grid tests are in use,-- monocular grid tests and stereoscopic grid tests. In the case of analogue instruments, the main objective of the monocular tests is to evaluate the metric performance of individual projectors in combination with their associated tracking and measuring devices. The measuring accuracies in the x and y directions are assessed through monocular measurement of grid points which have been projected from the negative plane into the instrument's projection plane. In the case of analytical instruments, where there is no physical reprojection of the rays, the monocular grid tests give the accuracy of the individual plates and carriages directly in terms of x and y image or photo coordinates. For both classes of instruments, the measuring accuracy in height is assessed using stereo grid tests. The overall accuracy in terms of X, Y and Z model coordinates is determined through the measurement of the grid model formed from the projection of a pair of grid plates.

Table 5.1 Hallert's test results for analogue instruments

Plotting machine	Machine No	L.H. Projector	R.H. Projector
Wild A7 (all C=150mm, Z=300mm)	No. 362	± 3.8um	± 4.6um
	No. 515	± 3.8um	± 4.4um
	No. 2153	± 2.8um	± 2.7um
Wild A8 (all C=150mm, Z=300mm)	No. 443	± 5.6um	± 8.3um
	No. 480	± 5.0um	± 5.4um
	No. 1443	± 4.9um	± 6.4um
	No. 2019	± 4.0um	± 3.2um
Zeiss Jena Stereometrograph (all C=150,Z=300mm)	No. 222320	± 4.1um	± 5.4um
	No. 222322	± 4.6um	± 3.1um
Zeiss Oberkochen Stereoplanigraph (C=152.5, Z=300mm)	No. 71133	± 3.3um	± 5.2um

Table 5.2 Savolainen's test results for analogue instruments

Machine	Machine No	m_x(um)	m_y(um)	m_z(‰ H)	Notes
Wild A8	No. 687	10.3	8.2	0.079	HUT, 1975
	No. 718	11.2	12.6	0.150	NBPRW, 1974
	No. 718	7.2	10.7	0.051	NBPRW, * 1975
	No. 936	9.4	12.0	0.076	NBPRW, 1974
	No. 338	4.3	3.7		Soil & Water,1974
	No. 460	11.4	6.4		Soil & Water,1974
	No. 2013	6.0	6.0		Soil & Water,1974
Wild A8	No. 703	5.8	11.2	0.096	NBS, 1971
	No. 535	--	--	0.106	NBS, 1974
	No. 638	--	--	0.060	NBS, 1974
	No. 686	--	--	0.081	NBS, 1974
	No. 703	--	--	0.096	NBS, 1974
Wild B8	No. 1626	--	--	0.069	HUT, 1973
	No. 1250	--	--	0.055	NBS, 1974
	No. 1254	--	--	0.076	NBS, 1974
	No. 1281	--	--	0.077	NBS, 1974
	No. 1284	--	--	0.120	NBS, 1974
	No. 1618	--	--	0.101	NBS, 1974
	No. 4692	--	--	0.081	NBS, 1974
Wild B8S	No. 4718	--	--	0.055	Finnmap, 1975
Zeiss Ober. Stereo-planigraph	No. 71552	5.6	4.9	0.070	HUT, 1972
	No. 71553	9.4	11.6	0.138	NBS, 1974

* These figures are the test results for Wild A8 with instrument No 718 and after repair and recalibration.

For **analogue plotters,** extensive testing took place in Scandinavia just over twenty years ago and the results were reported by Hallert et al in 1968. These figures are still of much interest and relevance today since so many of these types of analogue instruments are still in production use for high accuracy mapping and DTM work. Some of the results from Hallert's paper are listed in Table 5.1:

As stated above, these figures are still of interest since they show that the standard deviation values of $m_x=m_y=\pm 3$ to 5um for monocular grid tests of high precision analogue instruments have been realisable over a long period.

Some further tests of a similar nature have also been carried out by Savolainen & Ruotsalainen (1976) of the Helsinki University of Technology (HUT). They also quoted some results of additional tests carried out by other organisations in Finland such as the National Board of Survey (NBS) and the National Board of Public Roads and Waterways (NBPRW). Some results from these tests are quoted in Table 5.2.

Again these figures are of extreme interest in that, while the best figures agree with those of Hallert, there is also a wide variation in the accuracy figures for similar instruments. This shows that individual instruments may not have been serviced regularly nor checked for accuracy on a regular basis, resulting in a decided lowering of the accuracy available with a specific instrument.

Of course, many other tests of this type have been carried out; those quoted above are only examples. It is by no means the main purpose of this section to review all the test results. The only purpose of quoting these results is to give the reader some idea about the fundamental accuracy of these analogue photogrammetric instruments which have been used extensively to acquire DTM data.

From these results quoted above, it can be seen that it is very difficult to say what accuracy can be obtained from an individual analogue stereo-plotting machine, even though it is known which group (class, or order) it belongs to, because the exact status of the stereo-plotting machine in terms of its calibration is also a very important factor. For example, in some cases (see Table 5.2), the two best test results from the Wild B8 and B8S, which belong to the supposedly less accurate group of topographic plotters, are better than those from all but one of the several Wild A8 machines (which belong to the precision plotter class) and are even better than those from the Zeiss Oberkochen Stereoplanigraphs, which belong to the group of first order or universal machines.

Turning next to the accuracy of **analytical plotters** (APs), this is declared by manufacturers to be in the order of ±2-3um in x and y and ±6-7um in height, referred to the image planes. For example, in the case of the Zeiss Oberkochen Planicomp C-100, Hobbie (1977) stated that, "the absolute measurement accuracy in the photocarriage axes (and thus the image coordinates) is ±3um standard deviation over the entire photograph area." He also stated that a grid measurement carried out using 24 points gave RMS errors referred to the image plane of ±4um for x and y and ±6um for Z. In fact, the differences between the values declared by different manufacturers are very small. In practice, some of the APs give better results than the declared values. For example, Berg (1988) tested a Wild Aviolyt AC-1, and obtained a result of ±0.7 to 0.8um in x and y while the values declared by Wild are ±1.0 to 1.5um. Even Hallert testing an early OMI AP/C analytical plotter in 1968 obtained $m_x=\pm 1.6$um and $m_y=\pm 2.0$um.

However, it has also been found by Laiho and Kilpela (1988) that analytical plotters can behave in an unstable manner. They found that the coordinate readings may vary in a range of 2um to 5um between two successive measuring times and 3um to 10um over a whole day. Even though these figures are preliminary test results, they do give us some idea about the magnitude of the variation. The measuring results varied with time, the temperature, the movement of photo carriers, the position of the measured points, and even the hardware components of the instrument. Actually, the two authors also stated that it is still difficult at the present stage to draw conclusions about the reasons which cause these instabilities.

So just as with analogue instruments, different analytical instruments of the same type may exhibit different behaviour and different accuracy figures. Thus, it needs to be kept in mind that, in practice, the DTM data acquired by an analytical plotter may have been measured with an accuracy poorer than expected due to this instability.

5.3 The accuracy of the stereo-model

Having discussed the matter of the inherent accuracy of the photogrammetric instrument itself, it is now pertinent to discuss the accuracies which will be achieved when it is used with actual photographs. Ideally, after orientation, the resultant stereo-model should be an exact replica of the photographed terrain. But invariably this is not the case in practice due to various errors.

Stereo-model errors can be divided into two categories, i.e. the orientation errors and the non-orientation errors. The **orientation errors** include those errors which occur during interior, relative and

absolute orientation and may result from the mechanical and optical imperfections of the instruments or from errors or shortcomings on the part of the operator or from errors in the control points used for absolute orientation. The **non-orientation errors** are those errors due to atmospheric refraction, residual lens distortion, instability in the photographic material, etc. Generally speaking, the non-orientation errors are negligible with modern materials and equipment provided they are handled with care and attention.

The matter of the **inner orientation** should not give rise to serious problems provided, of course, that the various parameters -- the position of the principal point, the values of the camera focal length, lens distortion, etc -- have been accurately determined during camera calibration and that these values can be replicated accurately in the photogrammetric instrument.

The accuracy of **relative orientation** is affected by the quality of the photographs, the state of adjustment of the instrument, the skills of the operator, etc. For an analytical plotter, the accuracy of relative orientation should be much higher than that of analogue instruments since the instrumental accuracy does not depend on the accuracy with which the physical rotation of the plates may be made in the instrument.

The accuracy of **absolute orientation** is also affected by many factors such as the accuracy of the control points and their identification, as well as the accuracy of the instrument itself and the skills of the operator, etc. but it can be checked after orientation. Normally, as cited by Rinner and Burkhardt (1972), after orientation, there should be an average residual error in height of $\pm$0.05 - 0.10 per mil of the flying height (H).

5.4 The accuracy of height measurement

Since the accuracy of the measured elevation data is critical in the context of DTM accuracy assessment, so futher discussion may be justified. The accuracy of height measurement in static mode, m_h, which can be achieved by photogrammetric methods, can be expressed theoretically as follows:

$$m_h = H:B * H/f * m_p = S/B:H * m_p \qquad (5.1)$$

Where, m_p is the accuracy of parallax measurement; f is the camera focal length; H is the flying height; S is the reciprocal scale of the photography; and B:H is the base:height ratio.

This means that the height accuracy is dependent on the photo scale,

the base:height ratio and the accuracy of parallax measurement. From (5.1), it is obvious that, for a given camera with a certain focal length (f), the larger the photo scale leading to a reduction in the value of the flying height (H), the smaller the value of the scale factor, S, and the higher the accuracy of the measured DTM data. Therefore, the accuracy of photogrammetric measurement is directly related to the flying height. It becomes obvious that the larger the flying height, the smaller the photo scale, thus the poorer the accuracy of height measurement.

Of course, if the economic factor has to be considered, the scale of the aerial photography should be kept as small as possible, in which case, the flying height (H) would be greater. In terrain modelling practice, the photo scale and thus the flying height should be determined by the requirements of the DTM products, e.g. the accuracy of the DTM elevation values required for the specific project in question, or the scale and contour interval of the map which will be produced through the DTM, etc.

The base:height (B:H) ratio expresses the relationship of the distance between successive photographs and the flying height (H), which characterizes the geometry of the successive photographs and the resulting stereo-models. In general, the larger the B:H ratio, the more accurate the measurement will be. The B:H value is dependent on the focal length (f) and thus the angular coverage of the lens, the forward overlap as well as the photographic format, etc. If the forward overlap and the photo format are fixed, then the value of base:height ratio is determined by the value of the focal length of the aerial camera used. Thus the accuracy of height measurement is also determined by the camera focal length if other conditions, especially the scale, are all the same.

Stark (1976) carried out some experimental tests over the Rheidt test field in West Germany on the effect of the focal length of an aerial camera on the accuracy of height measurement. The same photo scale (1:10,000) was used in each case, which means that the flying height (H) was different for each flight. The test results conform to a general trend, but the values are not exactly the same as those given by theoretical models. Some of Stark's test results are given in Table 5.3.

Again, these results are of extreme interest, since the Planimat represents a very high precision analogue stereo-plotting machine and the PSK-2 is a high precision analytical instrument (stereocomparator). It is interesting to note that, in general, the PSK-2 gives somewhat better results in height than the Planimat, but really only significantly so in the case of the super-wide angle photography where

the Planimat results are inexplicably poor and quite out of line with the other results achieved with the instrument.

Table 5.3 Stark's test results for the effect of focal length on the accuracy of photogrammetrically measured data

Focal Length	Planimat				PSK - 2			
	m_x	m_y	m_Z		m_x	m_y	m_Z	
(mm)	(um)	(um)	(um)	(H%o)	(um)	(um)	(um)	(H%o)
305	4.8	5.0	21.1	0.069	6.7	5.2	16.6	0.054
210	5.8	7.8	11.0	0.052	5.0	6.3	12.5	0.060
153	5.8	7.4	13.3	0.074	3.4	4.1	8.2	0.054
85	4.6	6.7	12.0	0.141	4.2	4.9	6.3	0.074

Plotting out these results for the PSK-2 gives the general accuracy plot shown in Fig.5.1, where the theoretically expected results are also shown by solid lines.

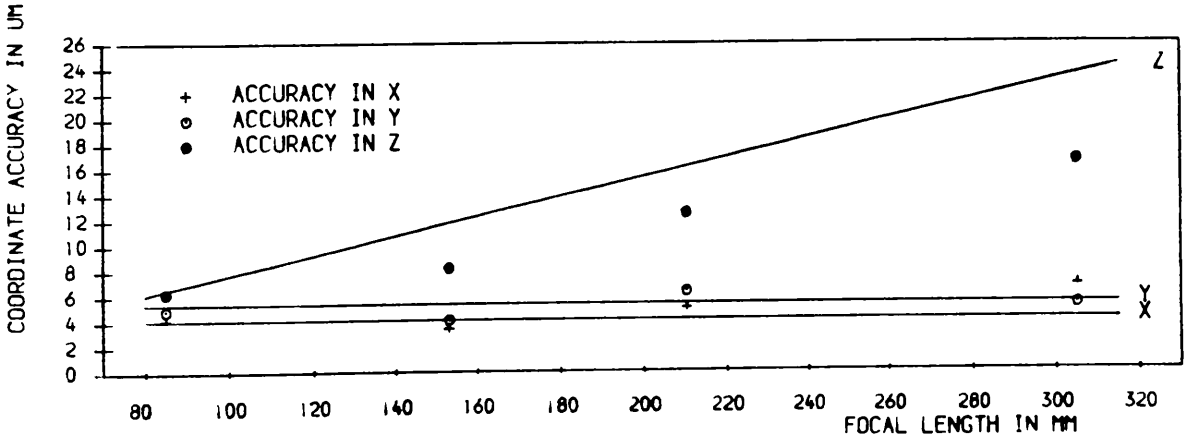


Fig.5.1 Accuracy of PMD varying with focal length

For planimetry, the standard deviation values both in x and in y are only varying slightly regardless of focal length and angular coverage with m_y values roughly 25 to 30% greater than the m_x values. With the **height** values, the accuracy does get better with increasing angular coverage and thus a corresponding increase in the base:height (B:H) ratio.

Of course, the accuracy of height measurement given in any test will also be affected by the accuracy of parallax measurement, which in turn is affected by many factors such as the photogrammetric instrument, the

skill of the operator, the accuracy of relative orientation, the resolution of the photographs, etc. These factors may affect the accuracy of measurement significantly. For example, some experimental tests on an operator's personal stereoscopic measuring capability (Schwarz, 1982) show that 70 to 90% of operators are measuring with a precision from 10 to 20um, which does seem to be unduly pessimistic. The discussion of the effect of these factors on the accuracy of parallax measurement is omitted here partly because such an effect is so complex that it is not easy to be discussed clearly in a few paragraphs and more importantly because what is mainly of concern is to obtain some idea about the overall accuracy of the photogrammetrically measured data, and not discuss in detail the sources of the errors in measurement.

5.5 The accuracy of statically measured photogrammetric data

Photogrammetric measurement can be carried in one of two modes, i.e. in either static or dynamic mode. The accuracy of data measured using these two different modes of measurement will be different since the accuracy with which the parallax measurement can be carried out is different in each case. In this section, some further results for static measurements will be given.

Gut & Hohle (1977) carried out extensive tests on the accuracy of photogrammetrically measured data from high altitude aerial photography. The photographs were taken at different times and with different focal lengths and different flying heights using a Wild RC-10 camera. The points were pre-marked with circular marks and quadrilateral pointers. The measurements were carried out on an analytical instrument - the Wild STK-1 stereocomparator - and two analogue instruments - both Wild A10 Autographs. Some of these results are listed below in Table 5.4.

Table 5.4 shows some astonishing results. In particular, the figures for height accuracy expressed in terms of per mil (‰) of the flying height (H) are quite an improvement on the already good figures produced by Stark. Indeed, they are some of the best figures produced by any form of testing, even better than many of those for grid tests given in Table 5.2. Partly, this is because, as stated before, the results listed in this table are obtained from pre-marked points on the photographs. But presumably the excellent flying and seeing conditions experienced over the Arizona Desert also have something to do with the superb results. In practice, with ordinary points and less good conditions for photography, the accuracy results will be much poorer.

Table 5.4 Gut's test results for the accuracy of photogrammetrically measured data from high altitude aerial photographs

Photo Scale	Flying Ht (H) (m)	Focal Length (f) (cm)	R. M. S. E					
			m_X (m)	m_Y (m)	m_H (m)	m_x (um)	m_y (um)	m_Z (‰ H)
1:33,000	5,000	15	0.24	0.39	0.13	7	12	0.026
1:33,000	5,000	15	0.23	0.37	0.13	7	11	0.026
1:56,500	5,000	8.8	0.75	0.54	0.36	13	10	0.072
1:58,000	8,800	15	--	--	0.37	--	--	0.042
1:57,000	8,400	15	--	--	0.30	--	--	0.036
1:58,000	8,900	15	--	--	0.33	--	--	0.037
1:58,000	8,900	15	0.66	0.70	--	11	12	0.034
1:88,000	13,400	15	0.54	0.92	0.52	6	11	0.039
1:63,000	13,400	21	0.54	0.82	0.78	9	13	0.060
1:78,000	11,900	15	--	--	0.41	--	--	0.035

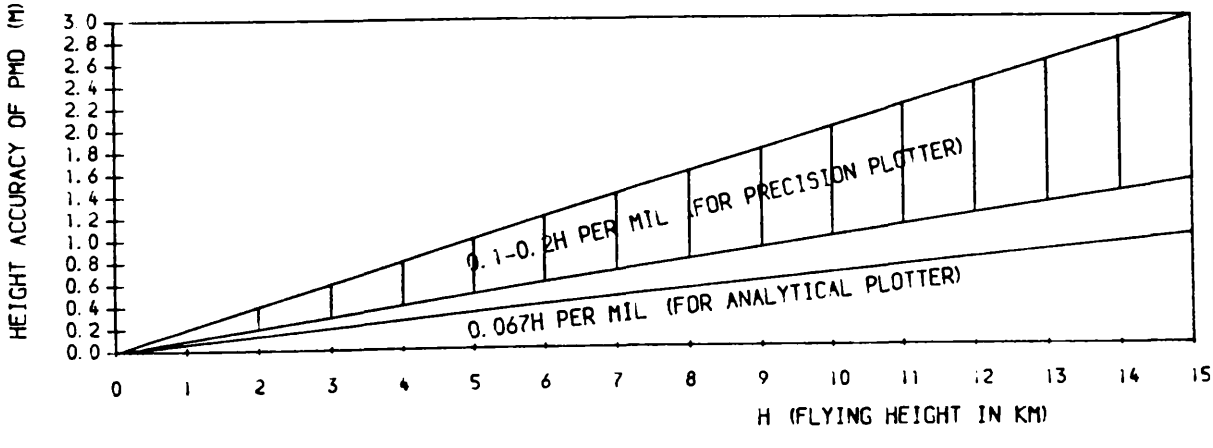


Fig.5.2 Accuracy of PMD varying with flying height (H)

Petrie (1987a) has summarized many of these test results and concluded that the following figures for spot height accuracy might be used for practical planning purpose:

Analytical Plotter: 0.067 per mil of the flying height (H)
Precision Plotter: 0.1 - 0.2 per mil of the flying height (H)

Actually, these are the values commonly accepted by most people in the

photogrammetric community as practical and realistic for the purposes of spot height accuracy in DTM work. Fig.5.2 shows the relationship (given above) between the height accuracy of statically measured data points and the range of flying heights over which the aerial photographs may be taken.

5.6 The accuracy of dynamically measured photogrammetric data

Dynamic measurement includes profiling and contouring. The accuracy of points measured dynamically in a stereo-model should be much lower than the values obtained in a static mode because there will be an additional error in the manipulation of the stereo-plotting machine by the human operator (keeping the floating mark on the stereo-model surface as it moves) and this will have a significant effect on the accuracy of the profile data. In the case of dynamic measurement, the speed of measurement is also a very important factor. In some cases, it also displays a systematic behaviour.

Alspaugh (1985) found that, in dynamic **profiling**, in addition to the factors discussed above, the direction of the profile relative to the model base and the direction (forward or reverse) in which a profile is traversed as the height is measured also affects the accuracy of measured points.

Sigle (1984) reported some tests on the accuracy of profile data in an area of Baden-Wurttemberg in West Germany. Wide-angle photography with photo scale of 1:30,000 was used. Six test areas with considerably different terrain forms were selected. Data points which had been measured statically with a Zeiss Oberkochen Planicomp C100 with an accuracy from 0.3m to 0.6m were used as check points. About 600 points from the profile data were compared with the check points. The mean height differences which resulted ranged from 1.5m (in flat terrain) up to 5m (in rough terrain) with a considerable systematic component of more than 2m for some test areas. If stated in terms of per mil of flying height, then the accuracy figures vary from 0.33 to 1.11 per mil of H with systematic errors of more than 0.44 per mil of H for some test areas. These results imply that an accuracy of 0.3 per mil of H is achievable but only if systematic errors are not too serious.

Photogrammetrically measured **contour data** is also a type of dynamically measured elevation data and the accuracy of this kind of data will again be lower than that of statically measured data and might be expected to be similar to that of profiled data. Indeed, quite independently, Rinner and Burkhardt (1972) suggested the use the same value of 0.3 per mil of H for the overall accuracy of photogrammetrically measured contour data, as has been mentioned above for

profile data.

5.7 Concluding remarks

In this chapter, the error sources likely to affect photogrammetrically measured data were sketched at the beginning. This account was then followed by a discussion of the accuracy of photogrammetric instruments, the accuracy of stereo-models and the accuracy of height measurement; In each case, some representative results of the tests carried out to establish the accuracy of photogrammetrically measured data have also been given.

Up to this stage, the three principal attributes of photogrammetrically produced DTM data, i.e. their distribution, density and accuracy, have all been discussed. This chapter will be followed by a discussion of the acquisition of DTM elevation data from the other main source -- existing contour maps.

Chapter 6

Data Acquisition from Existing Contour Maps

Chapter Six

Data Acquisition from Existing Contour Maps

6.1 Introduction

In Chapter 4, the acquisition of DTM data from aerial photographs using photogrammetric methods has been discussed and the accuracy of the resulting data (referred to as photogrammetrically measured data) has also been discussed in Chapter 5. In this chapter, the data acquisition from another data source - existing contour maps - will be discussed. First of all, a general discussion of the characteristics of contour maps will be made; then the existing methods of digitising maps will be reviewed; and finally, the accuracy of the digital contour data which is acquired from existing maps will be discussed.

6.2 Topographic Maps: Another main data source for terrain modelling

Every country possesses at least some topographic maps and these may be used as another main data source for (small-scale) digital terrain modelling. For many developing countries, this source may be poor due to lack of topographic map coverage or the deficient quality of the height and contour information contained in the map. However, for most developed countries, even some developing countries such as China, most of the terrain is covered by good quality topographic maps containing contours. Therefore, these form a rich source of data for terrain modelling provided that the limitations of extracting height data from contours are kept in mind.

6.2.1 General discussion of topographic maps

The scale and the contents of a basic topographic map series can vary from one country to another. Thus, for example, in the UK, the scale of the basic topographic map series covering the whole country is 1:10,000; in other countries (e.g. China) it may be 1:50,000. The classification of topographic maps according to scale and the degree of generalisation is shown in Table 6.1 (Konecny, 1979).

One important concern with topographic maps is the quality of the data contained in them, especially the metric quality, which is then specified in terms of accuracy. The fidelity of the terrain representation given by a contour map is largely determined by the density of contour lines and the accuracy of the contour lines themselves. One important measure of contour density is the **contour interval.** The commonly used contour intervals for different map scales are shown in Table 6.2 (Konecny, 1979).

Table 6.1 Different Scale Topographic Maps and Their Features

Map	Scale	Feature
Large to Medium Scale Topographic Base Maps	> 1:10,000	Representation true to plan
Medium to Small Scale Topographic Maps	1:20,000 to 1:75,000	Representation similar to plan
General Topographic Map	< 1:100,000	High degree of generalisation or signature representation

Table 6.2 Map Scales and Contour Intervals

Scale	Contour Interval
1 : 200,000	25 to 100 m
1 : 100,000	10 to 40 m
1 : 50,000	10 to 20 m
1 : 25,000	5 to 20 m
1 : 10,000	2.5 to 10 m

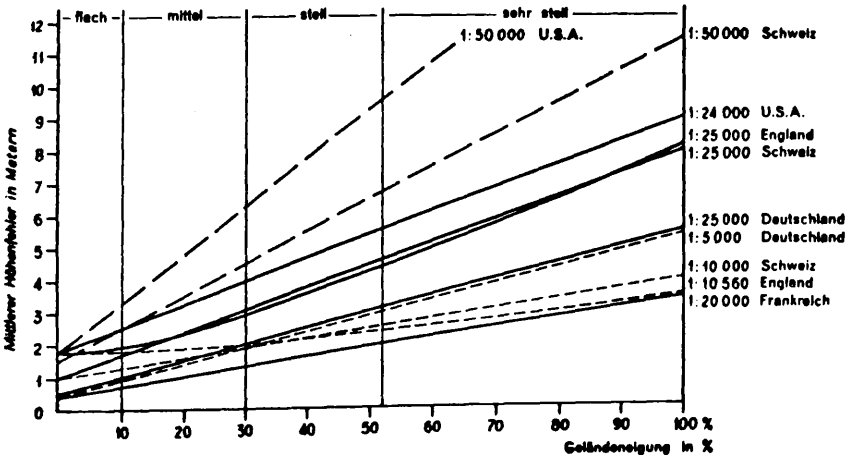


Fig.6.1 Examples of contour accuracy specifications (Imhof, 1965)

The accuracy requirements of a contour map are given by the so-called **map specification.** Examples of the specifications for the accuracy of contours for different map scales used in different countries are given in Table 6.3 (Imhof, 1965, Konecny, 1979) and some of them are shown diagrammatically in Fig.6.1.

Table 6.3 Contour Accuracy Specifications for Different Countries

Country	Scale	M_h in metres
W. Germany	1 : 5,000	0.4 + 3 tan A
Switzerland	1 : 10,000	1.0 + 3 tan A
Gt. Britain	1:10,000/1:10,560	$(1.8^2 + 3.0^2 \tan^2 A)^{1/2}$
Italy	1 : 25,000	1.8 + 12.5 tan A
NATO		3.0 + 7.5 tan A
Norway		2.5 + 7.5 tan A
Switzerland		1.0 + 7.5 tan A
Israel		1.5 + 5.0 tan A
W. Germany		0.8 + 5.0 tan A
Finland		1.5 + 3.0 tan A
Holland		0.3 + 4.0 tan A
Switzerland	1 : 50,000	1.5 + 10 tan A
United States		1.8 + 15 tan A
NATO		6.0 + 15 tan A

N.B. where A denotes the slope angle.

The approach for acquiring data from this source involves measurement of the contour lines contained in the existing topographic maps and is referred to as **cartographic digitisation.** Some kind of digitiser, e.g. a line-following digitiser or a raster scanner, is used to measure and transform the data from analogue form (on the map) to digital form. The digitised contour data is then used as the source from which the final DTM data is generated by some method of interpolation.

The accuracy of the digital terrain data acquired from contours on existing topographic maps is usually much lower than that acquired by ground survey or by photogrammetric methods. Therefore, topographic maps are primarily a data source for small-scale terrain modelling where accuracy demands are much lower.

6.2.2 Ordnance Survey (OS) Maps

To show an example of a topographic map series from which DTM data can be extracted, the various OS map series are briefly described. Another reason why such a discussion is included is that, in this project, a test on the accuracy of DTMs derived from an OS 1:63,360 scale topographic map has also been carried out (see Chapter 13). Some information about the topographic map series produced by OS GB is shown in Table 6.4 (Cole et al, 1983).

Table 6.4 Ordnance Survey Map Series

Map Scale	Ground Area (km)	Map Size (mm)	No. of Sheets in The Series	Contour Interval
1:10,000	5.0 x 5.0	679 x 559	10,198	5m or 10m
1:25,000	10 x 10	483 x 559	1,376	25 ft
1:50,000	40 x 40	1,000x890	204	50ft (in m)
1:63,360	40 x 45	1,000x725	198	50 ft

The height information of the last three series is simply derived from the basic 10,560 or 10,000 scale maps at which the contours have been surveyed, measured and compiled.

The 1:10,560 series is one of the oldest OS map series and it is the largest scale at which the whole island of Great Britain is covered. This series was the basic scale in mountainous and moorland areas. The contour interval is 25ft with a thicker gauge used to pick out the index contours at 100ft intervals. This series was replaced by the 1:10,000 series after metrication in early 1970's. The **1:10,000 series** covers the whole country with contour lines. Contours are shown at vertical intervals of 5m or 10m according to the relief. The contours are depicted on the map as thin continuous lines with a thicker gauge used for 25m (if a 5m interval is used), or 50m (if a 10m interval is used) to act as index contours.

The sheets in the older 1:10,560 scale series were produced originally in the 19th Century with selected contours surveyed by simple ground surveying methods and the intervening contours interpolated between them. Since the 1950s, the 1:10,560 (or 1:10,000) scale series has been completely resurveyed to a much higher standard using photogrammetric methods. For mountain and moorland areas, it was a complete resurvey of both the plan detail and the contours. However, for urban and rural areas, the planimetric information was derived from the larger scale maps (1:1,250 for urban areas and 1:2,500 for rural areas), but the contour information was measured separately and added to the derived planimetric map detail. Table 6.5 lists some information about the OS's resurvey work as applied to the measured contour data, all of which has been carried out using high precision plotting machines such as the Wild A8, Thompson-Watts Plotter, etc (Fagan, 1972).

The figures listed in this table can give fairly clear information about the quality of the contours which were measured in the photogrammetric resurvey. For example, if one takes the flying height (H) to be

3,800m and the possible contour interval to be H/1,500 (i.e. a C-factor of 1,500) which is quite feasible with instruments in the class of the Wild A8, then the minimum possible contour interval would be 3,800/1,500 =2.5m. Since the actual contour interval used by the O.S. is 5m or 10m according to the type of terrain, it can be seen that the O.S. solution is a quite conservative one and should lead to very good quality contour being available on the 1:10,000 scale maps. Looking at this matter from another viewpoint, if the accuracy figure of 0.3H per mil is accepted, for the flying height of 3,800m, this would lead to an accuracy (R.M.S.E.) of ±1.2m which would be very acceptable in relation to a 5m or 10m contour interval.

Table 6.5 Information about OS resurvey work

Area	Type of Survey	Photo Scale	Flying Ht	Camera
Mountain + Moorland	Resurvey	1:25,000	3,800m	W. A
	Continuous revision	1:25,000	3,800m	W. A
Urban + Rural	Contouring on sheets derived from larger scale maps	1:24,000 1:31,000	3,600m 2,700m	W. A S.W.A

The 1:25,000 scale series map sheets are simply photographically reduced from the 1:10.560 or 1:10,000 scale series. For those 1:25,000 scale maps which have been derived from the old 1:10,560 scale series, the contour interval is 25ft, while for those which have been derived from the newer 1:10,000 scale series, the contour interval remains 5m. Thus, the height information depicted on the 1:25,000 scale map sheets is the same as that shown on the corresponding 1:10,000 or 1:10,560 scale maps.

The vertical contour interval for the older **1:63,360 scale series** of maps is 50ft. Much of this information was in fact derived from the old surveyed contours rather than the newly contoured maps produced in the 1960s and 1970s by photogrammetric methods. After metrication, this series was replaced by a **1:50,000 scale series.** The First Series of 1:50,000 scale maps were basically a photographic enlargement of the 1:63,360 scale material. The contours shown on the First Series are those which appeared on the 1:63,360 scale series with the values simply converted from the actual foot values given on the 1:63,360 scale series (i.e. from the 50ft values) into the nearest equivalent values in terms of metres. The Second Series 1:50,000 scale maps produced more recently have contours which are derived from the photogrammetrically produced 1:10,000 scale series by photographic reduction with some generalisation followed by redrawing. The contours have a

10m vertical interval.

The relationship between these various Ordnance Survey map series of Great Britain is shown in Fig.6.2 (Chorley, 1987).

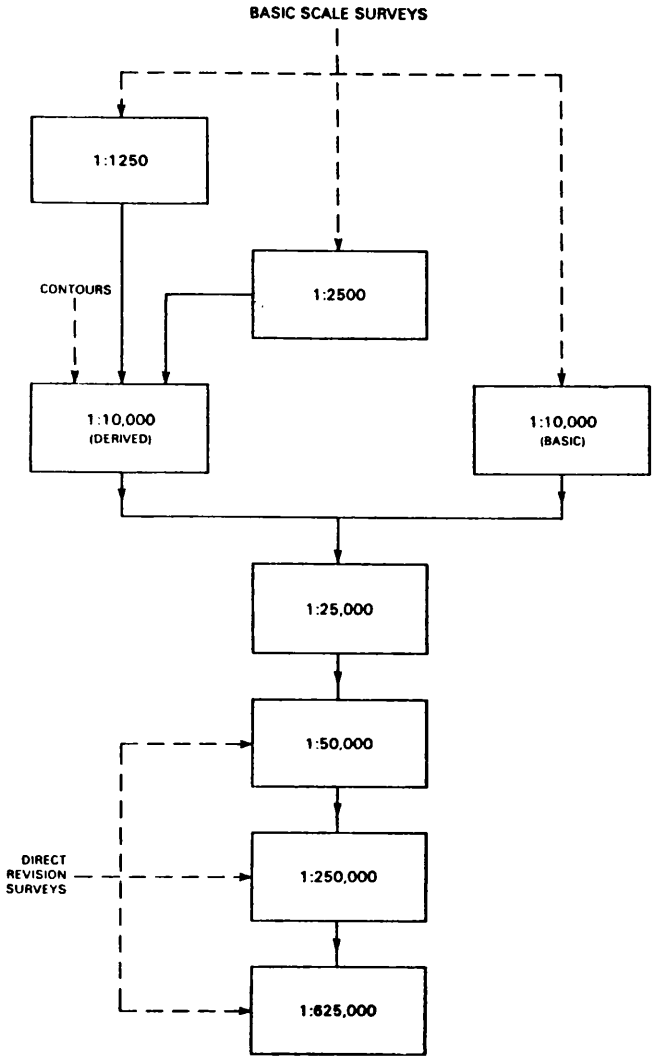


Fig.6.2 Links between OS map series (Chorley, 1987)

6.3 Accuracy of contour lines on a map

First of all, the accuracy of the contour lines shown on a map needs to be considered. Some discussion of this matter has already taken place in Section 5.6. In practice, a contour line can be regarded as consisting of a continuously connected string of points of a constant elevation and the accuracy of the contour line is a result of the errors present in this string of points. It is also well-known that the position and height of any point lying on the terrain surface can be defined by its 3-D coordinates (X, Y, Z) and there may be an error in each of these directions at this point. Usually, the two errors in

planimetry (i.e. in X and Y) are considered together as a vector, usually referred to as the **planimetric error** or **positional error.** Therefore, the overall accuracy of contour lines can also be specified by its accuracies in both planimetry and height.

The magnitudes of the deviations of a point lying on a contour from its true position (i.e. the planimetric error) may be measured, e.g. with respect to a reference contour which is produced usually at a much larger scale and using lower flying heights and high accuracy photogrammetric equipment. Some statistics may be constructed from these deviations, especially from the planimetric deviations. For example, the mean (average) values of the errors in the X and Y directions may be used as the measures of the systematic shifts (translations) of the contour lines in the X and Y directions and the ratio of these two mean values might be used as the measure of the mean direction error (rotation), etc. However, as to what kind of statistics should be constructed and used in practice all depends on the purpose of the accuracy assessment.

The height error and the positional error are also inter-related due to the existence of terrain slope. Koppe (1902, 1905) first described such a relationship using en empirical formula as follows:

$$m_C = m_h + m_{pl} \tan A \qquad (6.1)$$

Where, m_h refers to the accuracy of height measurement; m_{pl} is the positional or planimetric accuracy of the contour line; and A is the slope angle. Thus m_C is the overall accuracy of the contour in height including the effect of planimetric errors. This formula can also be shown diagrammatically as in Fig.6.3.

The overall planimetric accuracy of the contour lines, m_L, can then be determined by transposing the coefficients (m_h and m_{pl}) of Formula 6.1 and reversing the slope function as follows:

$$m_L = m_{pl} + m_h \cot A \qquad (6.2)$$

While the planimetric error in the contour alone is $m_{pl}=(m_C-m_h)\cot A$.

These formulae used together are usually referred to as the **Koppe Formulae** and are very widely used in the mapping communities in Central European countries such as Germany, Austria and Switzerland.

In this project, due to the difficulties which will be mentioned in Chapter 13, only the assessment of the height accuracy of DTMs will be considered. Therefore, what will be the particular concern of this discussion will be the height accuracy of digitised contour data.

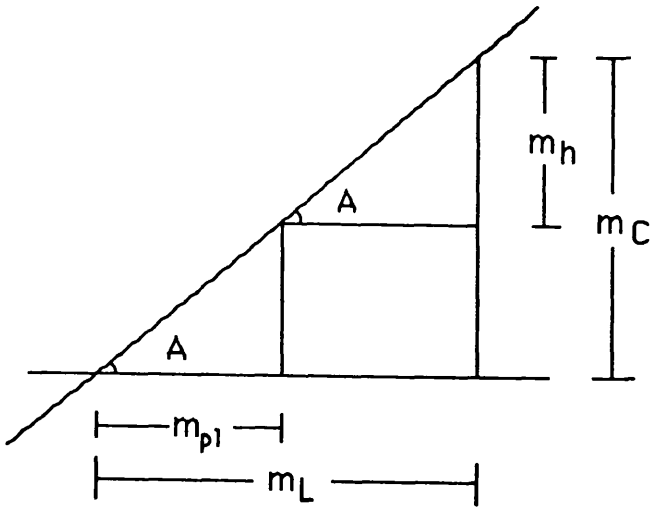


Fig. 6.3 Diagrammatic presentation of the Koppe Formulae

The accuracy of contour lines on a map can also be studied using one or other of two different approaches, i.e. by **theoretical analysis** or **experimental tests**, or by a combination of both these approaches. With the former, the errors which are likely to occur in different stages of the mapping process are estimated usually on the basis of extensive experience. Then the accumulated results are concatenated into an overall figure for contour accuracy. In the latter, as mentioned above, much more accurate contours, which have been produced from much larger scale stereo-models using much larger scale aerial photographs taken at much lower flying heights and employing high precision photogrammetric equipment, are used as reference (error-free) contours to check the deviation of those contours under investigation by superimposition. Then some measurements of the respective contours can be made and various statistics can be compiled and analysed to serve as the basis for estimating the accuracy of the contour lines.

Regarding this type of experimental investigation, a comprehensive study on the accuracy of contour lines was carried out many years ago by Lindig (1956). In his study, Lindig concentrated on the positional (or planimetric) errors. Indeed, the height errors are computed almost entirely from the corresponding positional errors. Lindig also introduced and constructed two additional errors, i.e. the **curvature error** and the **direction error** from the values of the positional errors. Fig.6.4 illustrates these errors diagrammatically considering points P_i' and P_{i-1}' lying on the original contour. In this diagram, H denotes the reference contour and H' denotes the contour under investigation. In this case, dv is the positional error; and dw is the direction error. The curvature error is defined as the difference betwen the curvatures at P_i and P_i'. The height error, dh, is computed from the positional error and is equal to dv.tanA, where A is the slope angle.

Lindig (1956) also devised some methods to measure the errors in position, direction and curvature directly from the superimposed contour lines. Lindig's work has been of considerable theoretical and practical interest and indeed, in spite of its limitations, it has been used subsequently by a number of researchers to give figures for contour accuracy.

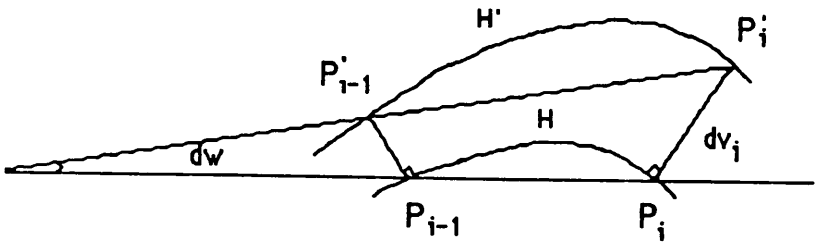


Fig.6.4 Contour errors defined by Lindig(1956)

In another later and largely theoretical study, Richardus (1973) carried out an investigation into the accuracy of the contour lines contained in large- and medium-scale topographic maps over the range of 1:250 to 1:50 000, where the angular slope values are assumed to vary from 4^{o} to 45^{o}. It was considered that the height accuracy of contour lines on the map is affected by the errors involved in the control points as well as in the photogrammetric measurements made in the stereo-model and the errors introduced by the overall mapping procedure.

Richardus also carried out an estimation of the loss in accuracy due to each of these factors mentioned above and produced a mathematical expression for the overall accuracy of contour lines on a map. However, Richardus's work has not been widely accepted or used in mapping practice. Therefore, no further or more detailed discussion is made here.

Through the analysis of the height accuracy achievable during contour measurement and all the associated positional errors which are likely to be encountered in the photogrammetric mapping process, it is also possible to make an overall estimate of contour accuracy. In the analysis carried out by Rinner and Burkhardt (1972), the average planimetric error for the points on contour lines was stated to be $\pm$0.2mm to 0.3mm at map scale. By taking $\pm$0.28mm as a representative value for the planimetric accuracy and $\pm$0.3 per mil of H for the height accuracy of contours on topographic maps, they produced some estimates, expressed in a form similar to that given by the Koppe Formulae, as shown in Table 6.6.

Table 6.6 Height accuracy of contours with map scale

Map Scale	Photo Scale	Focal(mm) Length	Flying Height (m)	Height Accuracy of Contours (m)
1: 5,000	1:15,000	153	2,300	0.7 + 1.4 tanA
1:10,000	1:20,000	153	3,100	0.9 + 2.6 tanA
1:25,000	1:30,000	153	4,600	1.4 + 5.8 tanA
1:50,000	1:60,000	115; 88	6,900;5,200	1.8 + 11. tanA *
1:100,000	1:70,000	115; 88	8,000;6,200	2.0 + 21. tanA *

* denotes that the average flying heights (6,000m and 7,000m respectively) were used for these estimates.

6.4 Cartographic digitisation methods

Unlike photogrammetric sampling, there are not many choices for cartographic digitising - one must carry out measurements of succesive positions lying along each contour line. The question of sampling is largely confined to making a decision about the intervals between successive positions measured along each contour. This section is going to review such limited methods.

The contours contained in a map can be digitised using one or other of two basic methods, i.e. either by line following or by raster scanning. The digitisation can also be carried out either as a manual operation or using automatic devices. Therefore, in essence, there are four possible ways to digitise a contour map. The characteristics of each of these methods have been fully discussed by Petrie (1987b) and are quoted in Table 6.7. From this table, it is obvious that manual raster digitisation is impractical. Therefore, there are only three methods in practical use.

The **manual line following digitising** method is still commonly used. In this method, either a mechanically-based digitising system or a solid-state digitising tablet can be used - nowadays usually the latter. In either case, the digitiser is operated manually. Contours are digitised either in point or stream mode, contour by contour. In **point mode**, the decision as to where each point should be measured is made by the operator. Usually, the actual measurement is made in a stationary (i.e. static) position to give the best possible accuracy. Where the **stream mode** is used, the tracing/measurement process is carried out in dynamic mode (i.e. with lower accuracy), with the coordinates being recorded

either on a time basis or on a distance basis. Although extremely common, manual digitising of contours is a tedious, demanding and time-consuming operation.

In order to overcome these difficulties encountered in manual digitisation, **automatic line-following devices** such as Laser-Scan's Fastrak and Lasertrak systems have been developed to remove the need for the manual movement of the cursor during measurement. However, an operator is still required to carry out supervision of the system and to execute various operations such as the initial positioning of the device on the contour; guiding the device through areas of close-packed contours, cliffs, etc; inserting contour elevation values, etc. Given this degree of manual intervention, the method is usually referred to as **semi-automatic digitisation.**

Table 6.7 Characteristics of contour digitising methods
(Petrie, 1987b)

		Line Following	Raster Scanning
Measurement		Points measured along contour lines only	The whole area of the map is scanned
Features		-Selective - only contour lines measured -Less data to be recorded and stored -Length of time required for measurement related to total length of contours	-Not selective - whole sheet scanned and measured -Very large amounts of data to be recorded and stored -Length of time required for measurement related to the size of sheet & the resolution of scan line
Characteristics	Manual Operation	-Low speed of measurement and data recording -Inexpensive hardware -Relatively easy feature coding	-Low speed -Enormous time required N.B. impractical to implement
	Automatic Operation	-Very high speed measurement and recording -Expensive hardware -Operator intervention for coding, etc.	-Very high speed measurement and recording -Very expensive hardware -Need for separate feature coding/labelling operation -Very considerable post-measurement processing required

Unfortunately, all automatic and semi-automatic line-following digitisers are very expensive and beyond the means of many smaller mapping organisations. Thus, the principal users are national topographic mapping agencies such as the Mapping & Charting

Establishment (UK), United States Geological Survey (USA), Geographic Survey Institute (Japan), Military Survey Directorate (Saudi Arabia), etc. and large private geophysical companies, e.g. GECO (Norway), and digitising bureaux, e.g.Taywood (UK), which have the capital resources to purchase, maintain and operate these systems.

The **raster scanner** makes fully automatic digitisation possible. In raster scanning, each line scan in the raster is divided into resolution units (pixels) of, say 25 to 50um, and, in each unit, the scan provides a return as to whether a contour line image is present or not. Each response may be recorded in some way, e.g. with a zero if nothing is present and a one if there is a line image. For contour digitising, very large format, high resolution raster-scanners are required. Both flatbed and drum-based designs are employed, but whichever is used, a huge capital investment is again required to purchase, maintain and operate such systems, e.g. those from Scitex, SysScan, Intergraph, Tektronix, etc.

As one can imagine, the data resulting from raster scanning is huge. For example, a map of 100cm square will produce four hundred million bits of information if the resolution is of the order of 50um, which is used in most cases. Furthermore, since most existing graphics software and hardware operate in vector mode, a raster-to-vector conversion is required both to make full use of existing technology and also to ensure a substantial degree of data compression.

6.5 Accuracy of digital contour data (DCD)

The term digital contour data can be used to refer to two types of data. One is the data directly measured and recorded from a stereo-model while the original contouring is in progress. While the other is the data which is digitised from the contour lines shown on published hard-copy maps. These two types of data will have different accuracies because the latter has undergone various cartographic and digitisation processes. The accuracy of photogrammetrically measured contoured data has already been discussed in the previous chapter. In this section, only the accuracy of the digitised hard-copy contour data will be discussed.

As explained above, the manual method is still the dominant one used for contour digitising. Therefore, the discussion will be confined to this type of digitising.

As was the case with data acquisition using photogrammetric methods, the accuracy of digitised data is also the overall result of the errors from different sources, including the basic accuracy of the contours

themselves, the accuracy of the digitiser, and the fidelity with which the operator carries out measurement of the contours. The accuracy of the contour lines shown on the contour maps has been discussed before, so only the accuracies of the digitiser and of the measurement process will be given here.

First of all, there is a matter of the inherent **resolution** (or least count) given by the digitiser itself, which is a very important factor in indicating the potential accuracy of a digitiser. The absolute accuracy of the digitiser will of course be lower than the resolution. Some examples of such information are quoted in Table 6.8 from Kelk (1973).

Table 6.8 Accuracy of digitisers

Device / Parameters	Mechanically-Based Pencil-Follower	Solid State Tablet
Resolution (least count)	0.025mm to 0.1mm	0.025 mm
Repeatability (precision)	±0.01mm to 0.1mm	±0.025 mm
Absolute positional static accuracy	Better than ±0.1mm to 0.15mm	±0.15mm/m
Dynamic accuracy	<0.2mm up to 2.5cm/sec.	

Table 6.9 Rollin's Test Results on Digitisers

Digitising Table	No.of Tables Tested	Quoted Accuracy	RMSE
Ferranti Freescan	23	±0.127mm	±0.088mm
Altek Datatab 3	8	±0.076mm	±0.075mm
Altek Datatab 2	1	±0.127mm	±0.090mm
Aristo Aristogrid 100	1	±0.127mm	±0.142mm
Kontron Digikon 2436	1	±0.10mm	±0.087mm

Also like stereo-photogrammetric instruments, the **accuracy** of an individual digitiser may also vary with its condition in terms of its maintenance. Rollin (1986) reported some figures from the calibration of the manually-operated tablet digitisers used by the OS, and found that the poorest results came from one of the older Ferranti Freescan solid state tablets which displayed an RMSE in x and y of ±0.123mm and

±0.168mm respectively. But in general, most of the RMSE values lie in the range from ±75um to ±90um which may be regarded as very satisfactory having regarded to the intrinsic accuracy (±0.2 to 0.3mm) of the contours being measured. Some results from Rollin's tests are given in Table 6.9.

When manual digitisation is carried out, an additional error is introduced when using the cursor of the digitiser to measure the contour lines. For such errors, an accuracy of ±0.10mm in terms of R.M.S.E at map scale is a reasonable estimate because the resolution of the human eye is believed to be around such an amount. The overall accuracy of digitising might be up to ±0.15mm. This value may be appropriate for the measurement of contours in **point mode** which corresponds to the static mode of the measurement on stereo-plotting machines. However, like the dynamic mode of measurement in a photogrammetric stereo-model, the accuracy of the measurement of contours in **stream mode** will also be lower, though it is difficult to tell how much lower. An estimate based on experience indicates that a figure of ±0.2 to 0.25mm might not be inappropriate.

In some organisations, the digitising process is carried out on an enlarged copy of the map, so that better conditions will be created for the operator's measurement. Therefore, the accuracy of digitising should be rather better than some of the values quoted above. In the case of the OS, the original document is usually enlarged by a factor of 1.5x. In this case, the accuracy figure of ±0.15mm at the enlarged map scale is equivalent to only ±0.1mm at the original map scale. Taking into account the improvement in the conditions for measurement, then the accuracy may be better than ±0.1mm. Sometimes, this is, of course, not such a serious problem since the positional error for contours allowable by the map specification is often larger than this value.

6.6 Concluding remarks

In this chapter, some discussion of the characteristics of contour maps has been made, including their specified accuracy. In particular, the OS map series has been described as an example. Then, the various methods of contour digitising and their characteristics have been reviewed; and the height accuracy of the contour data digitised from existing contour maps has also been discussed. Finally, a brief survey of the accuracy of digitisers is also given, followed by a discussion of the accuracy of digitising and digitised data.

Up to this stage, the main topics concerning DTM data acquisition have been discussed and the accuracies of the raw (source) data from different sources have been examined in some detail. However, since the

discussion of the accuracy of the source data has, till now, only been related to random error, some more discussions about gross error and its detection and the improvement of the quality and accuracy of the source data with some pre-processing procedures are also desirable. It might be also reasonable to allocate the chapter about these discussions (on data pre-processing) in the next chapter. However, it will be introduced as Chapter 8 since the matter of the procedures used for surface reconstruction and the characteristics of the DTM surface itself are also extremely important in this context. Therefore these matters will be covered in the chapter which follows in order to finish off the discussion about theory and background to the the present project, and thus, to complete the first part of this thesis.

Chapter 7

Digital Terrain Surface Modelling

Chapter Seven

Digital Terrain Surface Modelling

7.1 Introduction

As discussed in the introductory chapter, the main factors affecting the accuracy of final digital terrain models are the type of terrain itself; the attributes of the source data (including its accuracy, pattern and density); and the type of terrain model surface required - which is linked to the techniques used for surface modelling. In the previous chapters, the types of terrain surface have been described in Chapter 3; the attributes of photogrammetrically measured data have been discussed in Chapters 4 and 5; and the characteristics of the digital contour data have also been discussed in the Chapter 6. However, the matter of the techniques used for modelling the terrain surface still remain to be discussed before the assessment of DTM accuracy can be carried out.

In this chapter, a discussion of various topics related to the modelling of the terrain model surface from the data sets which can be obtained by both photogrammetric sampling and cartographic digitisation will be undertaken. First of all, the basic concepts will be introduced and defined; the various approaches which can be used for terrain surface modelling will outlined; and the types of terrain model surfaces which can be obtained from data sets with different patterns and by different modelling approaches will also discussed; From these discussions, the importance of different types of terrain model surface to the present study can then be judged and it can then be decided which types of terrain model surface should be taken into serious consideration when the DTM accuracy assessment is carried out later in this project.

7.2 Interpolation and surface reconstruction

7.2.1 Basic concepts

A digital terrain model (DTM) is a mathematical (or numerical) model of terrain surface. It employs one or more mathematical functions to represent the terrain surface according to some specific methods based on the set of measured data points. The mathematical functions are usually referred to as **interpolation functions.** The process by which the representation of the terrain surface is achieved is referred to as **surface reconstruction** or **surface modelling** and in the published literature, the actual reconstructed surface is often referred to as the **DTM surface.** Therefore, terrain surface reconstruction can also be

considered as **DTM surface construction** or **DTM surface generation.** After this reconstruction has been achieved, the required height information for any point in the model can be extracted from the DTM surface.

Interpolation is a frequently discussed topic in the area of digital terrain modelling. A lot of work has been done by many investigators. A comprehensive review of the interpolation methods used in DTM work has been given by Schut (1976). In Schut's paper, six groups of interpolation methods were distinguished, namely, (i) moving surface; (ii) summation of surfaces; (iii) simultaneous patchwise polynomials; (iv) interpolation in a rectangular grid; (v) interpolation in a net of triangulation; and (vi) interpolation in a string DTM.

From different viewpoints, these methods may be classified in different ways. For example, according to Wild (1980), the following four main groups can be distinguished instead, namely, (i) polynomial interpolation; (ii) least squares prediction; (iii) summation of surfaces; and (iv) interpolation with finite elements.

These classifications are more or less based on the characteristics of the interpolation methods themselves. Some other alternatives can also be used. For example, Petrie and Kennie (1986) classified these methods into three groups under the heading of pointwise, patchwise and global interpolation, based on the size of the individual areas used to represent the DTM surface. This classification was made in the context of random-to-grid interpolation procedures. In addition, triangulation was treated by them as a separate method since interpolation to form a basic network of elevations is not a requirement to form the DTM surface.

The concept of interpolation is a little different to that of surface reconstruction. The former includes the whole process of estimating the elevation values of new points which may in turn be used for surface reconstruction, while the latter emphasises the process of actually reconstructing the surface, which, as noted above with the triangulation method, may not involve interpolation. To attempt to clarify this matter further, **surface reconstruction** only covers those topics concerned with "how the surface is reconstructed and what kind of surface will be constructed", e.g. is it a continuous curving surface or does it consists of a linked series of planar facets.

By contrast, **interpolation** covers a much wider range. It may include the matter of surface reconstruction and that of extracting height information from the reconstructed surface; it may also include the formation of contours either from randomly located points or by from a measured set of elevation values obtained in a regular gridded pattern. In both of these latter cases, the measured values are honoured in the

resulting surface and the interpolation process only takes place after the surface reconstruction, either to extract height information for specific points or to construct contoured plots.

In this chapter, various topics concerning the characteristics of DTM surfaces and the construction (or generation) of such surfaces are those of most concern. First of all, the approaches for constructing DTM surfaces from different data patterns will be discussed in the following section; then the different types of DTM surfaces which can possibly be constructed from a data set will be briefly sketched in the section after that. Finally, for the sake of completeness, some brief discussions of the methods (procedures or algorithms) used in these approaches to form special data patterns will also be presented.

7.2.2 General polynomial function for terrain surface realization and interpolation

Before starting to discuss the various approaches for DTM surface construction (or generation), it seems pertinent first to introduce the mathematical function widely used in DTM practice to realize these DTM surfaces.

As has been discussed in Chapter 2, the mathematical expression for a DTM surface can be defined explicitly as follows:

$$Z = f(X,Y) \qquad (7.1)$$

The most widely used function for realization of this expression is the polynomial function as shown in Table 7.1.

Table 7.1 Polynomial function used for surface reconstruction
(Petrie & Kennie, 1986)

Individual Terms	Order	Descriptive Terms	No.of Terms
$Z= a_0$	Zero	Planar	1
$+ a_1X + a_2Y$	First	Linear	2
$+ a_3XY+ a_4X^2 + a_5Y^2$	Second	Quadratic	3
$+ a_6X^3 + a_7Y^3 + a_8X^2Y + a_9XY^2$	Third	Cubic	4
$+ a_{10}X^4+ a_{11}Y^4+ a_{12}X^3Y+ a_{13}X^2Y^2 + a_{14}XY^3$	Fourth	Quartic	5
$+ a_{15}X^5+ \quad \ldots$	Fifth	Quintic	6

The characteristics of each term in this function have been fully discussed by Petrie and Kennie (1986) and are shown graphically in Fig.7.1.

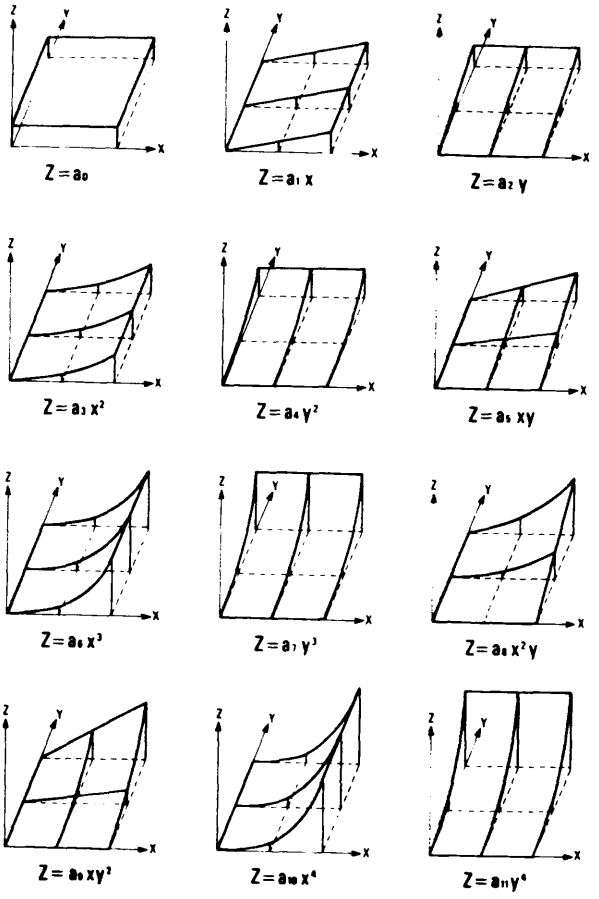


Fig.7.1 Surface shapes produced by individual terms (Petrie & Kennie, 1986)

For the generation of the actual surface in a specific modelling program, it is not necesssary (and in practical terms, impossible) to use all the terms inherent in this function. In practice, only a few terms are used, the selection of these being decided upon by the system designer and implementor. Only in a very few cases is there the possibility for the user to select which terms in the function might be most appropriate for modelling the specific piece of terrain in question.

As have been shown in Fig.7.1, each individual term of the general polynomial function has its own characteristics in terms of shape. By using certain of these terms, a surface with special characteristics can be constructed.

7.3 Alternative approaches for digital surface modelling

As discussed in Chapter 4, various data patterns are available using different sampling methods for DTM data acquisition. From these different data sets, different approaches can be used for DTM surface construction and different types of surfaces can also be constructed. However, a general discussion of the various types of DTM surfaces will be given in the next section. In this section, only the approaches for constructing these surfaces from different data patterns themselves will be examined.

Three main approaches can be distinguished for modelling the terrain surface, namely, point-based modelling, triangle-based modelling and grid-based modelling. Also a hybrid approach combining any two of these is also possible.

7.3.1 Point-based surface modelling

If the **zero order** term in the polynomial is used for DTM surface realization, then the result is a horizontal (or level) plane, as shown in Fig.7.1. At every point, a horizontal (or level) planar surface can be constructed. If the planar surface constructed from an individual data point is used to represent the small area around the data point (also referred to as the region of influence of this point in the context of geographical analysis), then the whole DTM surface can be formed by a series of such contiguous horizontal planar surfaces. The resulting overall surface will be discontinuous.

For each individual horizontal planar sub-surface, the mathematical expression is simply as follows:

$$Z_i = H_i \qquad (7.2)$$

Where Z_i is the height on the level plane surface for an area around point I and H_i is the height value of point I.

This approach is very simple. The only difficult thing which needs to be done is to define the boundaries between the adjacent areas. As stated before, the discussion about this type of surface and the determination of the area boundaries will not be given here but later in Section 7.4.1. Since this approach forms a series of sub-surfaces based on the height information of individual points, so the modelling based on this approach can be regarded as **point-based surface modelling.**

Theoretically, this approach is suitable for any data pattern since it concerns only individual points. Furthermore, irregularly shaped areas

(and planar surfaces) based on irregularly distributed data points are also possible. However, as far as the process of determining the boundaries of the region of influence by each point is concerned, the computation will be much simpler if regular patterns such as a square grid, equilateral triangle, hexagon, etc are used. Although it would seem quite feasible to implement this approach in surface modelling, it is not really a practical one due to the resulting discontinuities in its surface as will be discussed later.

7.3.2 Triangle-based surface modelling

If more terms are used, then a more complex surface can be constructed. Inspection of the first **three terms** (the two first order terms together with the zero-order in Table 7.1) shows that they form linear surfaces. To determine the three coefficients of this particular polynomial, three data points are the minimum requirement. These three points can form a spatial triangle, in which case, a tilted planar surface can be defined and constructed.

If the plane surface determined by an individual triangle is used to represent only that area covered by the triangle, then the whole of the DTM surface can be formed by using a linked series of contiguous plane triangles. Again, the discussion about the resulting surface will be given in Section 7.4.2. The modelling based this approach is usually referred to as **triangle-based surface modelling.**

The triangle may be regarded as the most basic unit in all geometrical patterns in any case, since a regular grid of square or rectangular cells or any polygon with any shape can be decomposed into a series of triangles. Therefore, triangle-based surface modelling is the approach which is feasible with any data pattern no matter whether it has resulted from selective sampling, composite sampling, regular grid sampling, profiling or contouring. Since triangles have a great flexibility in terms of their shape and size so this approach can also easily incorporate break lines, form lines, discontinuities and any other data. Therefore, the triangle-based approach has received increasing attention in terrain modelling practice, and can be regarded as one of the main approaches that can be taken in surface modelling.

The specific manner in which the triangle-based modelling approach is actually implemented in modelling practice is not the main concern of this research project. Nevertheless, in view of its current importance in practice, a brief discussion about the methods of forming triangular networks will be given in Section 7.5.

7.3.3 Grid-based surface modelling

As discussed above, the use of the first three terms in the general polynomial to generate spatial triangles is very feasible and is becoming widely used in practice. However, if more than the first three terms of the polynomial are used, then the use of other approaches may be considered although, as will be seen later, the additional terms can also be used with triangles to create curved facets if a linked series of triangles forming a polygon is being considered.

If the first three terms, together with the term a_3XY of the general polynomial, are used for DTM surface construction, then four data points are the minimum requirement to form a surface in this case. The resulting surface is referred to as a **bilinear surface.** Theoretically, quadrilaterals of any shape such as a parallelogram, rectangle, square, or an irregular one can be used as the basis for this type of surface. However, for practical reasons, such as the resulting data structure and the final surface presentation, a regular square grid is the most suitable pattern. As in the case of triangle-based surface modelling, the resulting surface will consist of a series of contiguous bilinear surfaces.

Higher-order polynomials can also be used for DTM surface construction. The use of a high-order polynomial function to construct a DTM surface for a large area may lead to unpredictable oscillations in the resulting surface. In order to reduce the risk of this situation arising, a restricted number of terms - usually only the second- and third-order terms - will be used in practice. The minimum number of elevation points which will be necessary to construct the DTM surface will of course be governed by the number of terms used, but in any case, the number will be greater than four. In this case, different patterns and geometric figures other than the basic triangle or square grid cell can be considered for use in the surface reconstruction. Nevertheless, because of the difficulties likely to be encountered in data structuring and handling, elevation source data which is evenly distributed, as in the case of regular grid and equilateral triangle patterns, is still of significant importance.

In fact, from the practical point of view, as has been discussed in Chapter 4, grid data has many advantages in terms of data handling. Therefore, it might be said that elevation grid data from regular grid sampling and progressive sampling, especially the square grid data, is particularly suitable in this case. Indeed, such are the advantages that many DTM packages only accept gridded data for surface construction, in which case, a preliminary data pre-processing operation (random-to-grid interpolation) is necessary to ensure that the input data is in the required form.

The approach of surface modelling based on a regular grid is usually referred to as **grid-based surface modelling.** In practice, this approach to surface modelling can be used to construct a smooth surface and is often applied to a global data set covering rolling terrain. It will be obvious that it has less relevance (or application) to broken terrain with steep slopes, numerous break lines, sharp terrain discontinuities, etc. A further discussion about the surfaces resulting from the grid-based approach will also be given in the next section.

7.3.4 Hybrid surface modelling

The actual data structure which is implemented using a particular geometric pattern for surface modelling is usually referred to as a **network** in the context of terrain modelling. So, on this basis, it can also be said that a DTM surface is usually constructed from one or the other of two main types of network - a gridded network or a triangular network. However, a **hybrid approach** has also been widely used to construct DTM surfaces. For example, a gridded network may be broken down into a triangular network to form a contiguous surface of linear facets. Going in the opposite direction, a gridded network may also be formed by interpolation within an irregular triangular network.

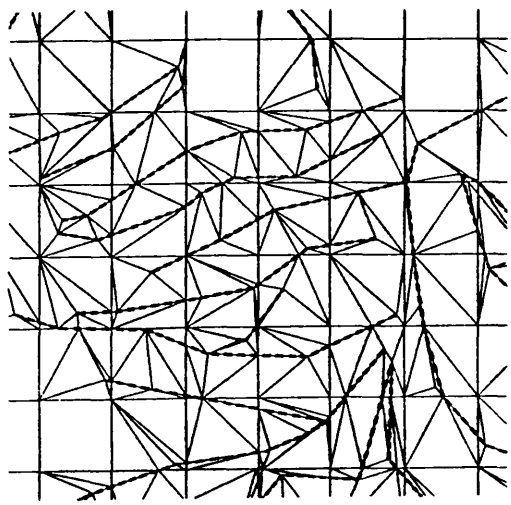


Fig.7.2 An example of the hybrid approach to surface modelling (HIFI)

One of the latest developments in this hybrid approach, which is now being used in the HIFI 88 and CIP packages, is to have a basic grid of squares or triangles obtained by systematic grid sampling. Where break lines and form lines are available (as in the case of composite sampling), the regular grid is broken into and a local irregular triangular network is implemented. Fig.7.2 is an example produced from the HIFI package showing the hybrid approach to surface modelling.

It might be also possible to combine point-based modelling with either grid-based modelling or triangle-based modelling to form a hybrid approach. That is, the boundaries of the region of influence of a point can be determined using either a gridded or a triangular network in the case of data located in a regular pattern or based on a triangular network in the case of irregularly located data. Examples are shown in Fig.7.3 (Section 7.4.1). In the case of Fig.7.3(a), the surface has been constructed from regular data in a square grid using point-based modelling; while the surface shown in Fig.7.3(b) has been constructed from a regular triangular network also using point-based modelling.

7.3.5 Alternative methods for surface modelling

Attempting to summarize the discussion up to this point, there are, in principle, three possible approaches for surface modelling - point-based, triangle-based and grid-based approaches. Each is suitable for a specific data pattern. However, in practice, point-based modelling is not practical and is therefore not widely used. Only the triangle-based and grid-based approaches to surface reconstruction have been widely used and so these were considered as the two basic approaches by Petrie and Kennie (1986).

In fact, as Petrie and Kennie (1986) have emphasized, there are also **two basic methods** which have been used for a digital terrain model (DTM) surface construction in these modelling approaches as follows:

i). direct construction from measured elevation data; or
ii). indirect construction from derived data.

Thus, the surface can be constructed directly from the **original source data**, e.g. by using square grids or regular triangles or through triangulation in the case of randomly located data. Also as noted above, a hybrid approach combining grids and triangles or a mixture of equilateral and irregularly shaped triangles can be employed with the measured source data. Alternatively, the surface can be constructed from elevation data which has been **derived by interpolation** from the original measured data, as is the case when a preliminary random-to-grid interpolation has been carried out before the DTM surface construction is attempted.

The above discussion shows that various different approaches can be used to construct different types of DTM surfaces even from the same data set. Another point which needs to be made here is that the different approaches to surface modelling may be classified further based on different points of view, so providing a further insight into the surface reconstruction process.

7.4 Classification of DTM surfaces

DTM surfaces can be classified based on different criteria. For example, they may simply be classified into three groups based on the **size** and coverage of the areas which are represented by the individual elements which go to make up a DTM surface. On this basis, one can define either:

i). a local surface;
ii). a regional surface; or
iii). a global surface.

The usual basis on which a **global surface** may be constructed is a situation comprising a very large data set and covering a large area of terrain containing very simple and regular terrain features. Alternatively it may be used when only very general information about the terrain surface is needed for purposes such as reconnaissance. By contrast, the construction of **local surfaces** could be based on the premise that the area to be modelled is very complex or that only a local area is of interest. The reasons for using a **regional surface** could be the result of some type of compromise between those criteria given for using a global surface and those used to justify the use of a local surface.

Another criterion which may be useful in the classification of DTM surface models is to make use of certain specific characteristics of these models. One of the most important characteristics of a DTM surface concerns the **continuity** of its surface. Based on this criterion, DTM surfaces can be classified further into three groups, namely,

i). those with discontinuous surfaces;
ii). those exhibiting a continuous surface; and
iii). those having a smooth surface.

7.4.1 Discontinuous surfaces

A **discontinuous surface** is the result of the thought that the height value of any measured point is representative for the values in its neighbourhood (Peucker, 1972). On this basis, the height of the point to be interpolated can be approximated by adopting the height of the closest reference point. In this way, a series of local level surfaces are used to represent the terrain surface. Theoretically speaking, this type of surface is the realization of the so-called point-based surface modelling which has been discussed before in Section 7.3.1.

As has been pointed out in the discussion of point-based surface

modelling, this type of surface can be constructed from any type of data set, no matter whether it is regular or irregular.

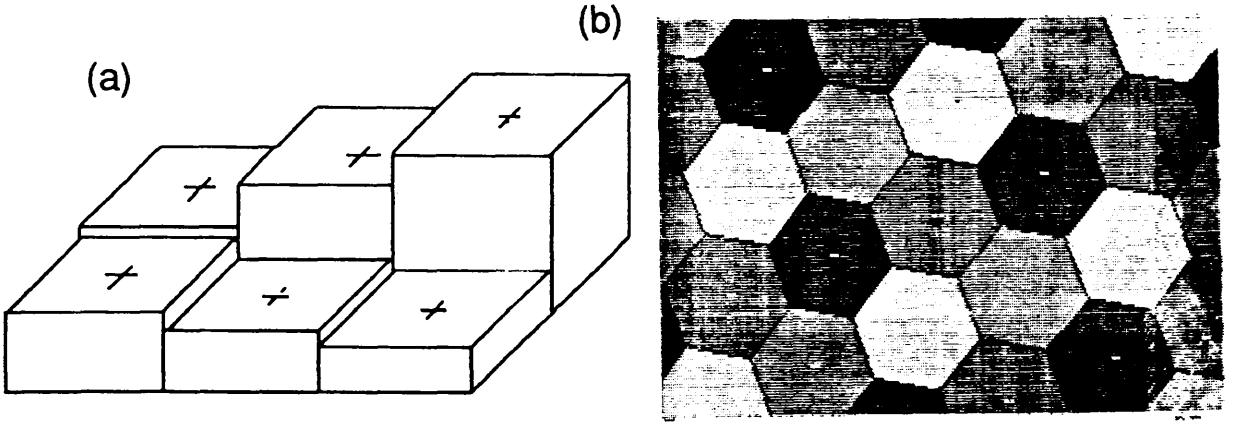


Fig.7.3 Discontinuous surface constructed from different patterns (a). from regular gridded data; (b). from equilateral triangular data.

For the regular data, the determination of boundaries between the sub-surfaces is much easier. Fig.7.3(a) shows such a surface constructed from gridded data sets, while Fig.7.3(b) shows the plan of a discontinuous surface consisting of a series of horizontal planes each with a hexagonal shape resulting from a data set with an equilateral triangle pattern (Peucker, 1972). From these diagrams, it can be seen clearly how the boundaries are determined in each of these cases. Therefore, no further explanation needs to be given here.

However, whenever the data is irregularly distributed, then the boundaries of the so-called region of influence of each of the data points needs to be determined. Normally, this is done by constructing the so-called Thiessen polygons which are widely used in the context of geographical analysis since this method was proposed by the climatologist A. H. Thiessen in 1911 (Brassel & Reif, 1979). Actually, the Thiessen polygon is a region enclosed by a embedded series of perpendicular bisectors, each located midway between the point which is under consideration and each of the surrounding neighbours. Fig.7.4 shows such a polygon.

Again, a level plane can be constructed over the area enclosed by a Thiessen polygon. Therefore, a whole series of discontinuous surfaces can be constructed from the network of Thiessen polygons derived from the data set. In the literature, such a network is referred to as the Thiessen diagram, or Voronoi diagram, or as Wigner-Seitz cells or the Dirichlet tesselation - the actual term used seeming to vary between different scientific disciplines, although the basic idea is common to them all.

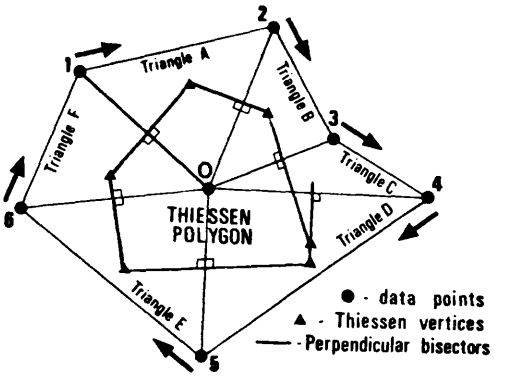


Fig.7.4 Thiessen polygons and Delaunay triangles (Petrie and Kennie, 1986)

It can also be seen from Fig.7.4, that the dual of a Dirichlet tesselation (or Thiessen Diagram) is a triangulation. This dual relationship was first recognized by Delaunay (1934). Therefore, such a triangulation is usually named after Delaunay. Fig.7.5(a) and (b) show a Thiessen diagram and its dual - the corresponding Delaunay triangulation. Nowadays, the actual procedure to determine the Dirichlet tesselation (or Thiessen diagram) is to carry out first the Delaunay triangulation, then to define the Thiessen polygons from the triangular network. More discussion about the Delaunay triangulation will be given later in this chapter in another context.

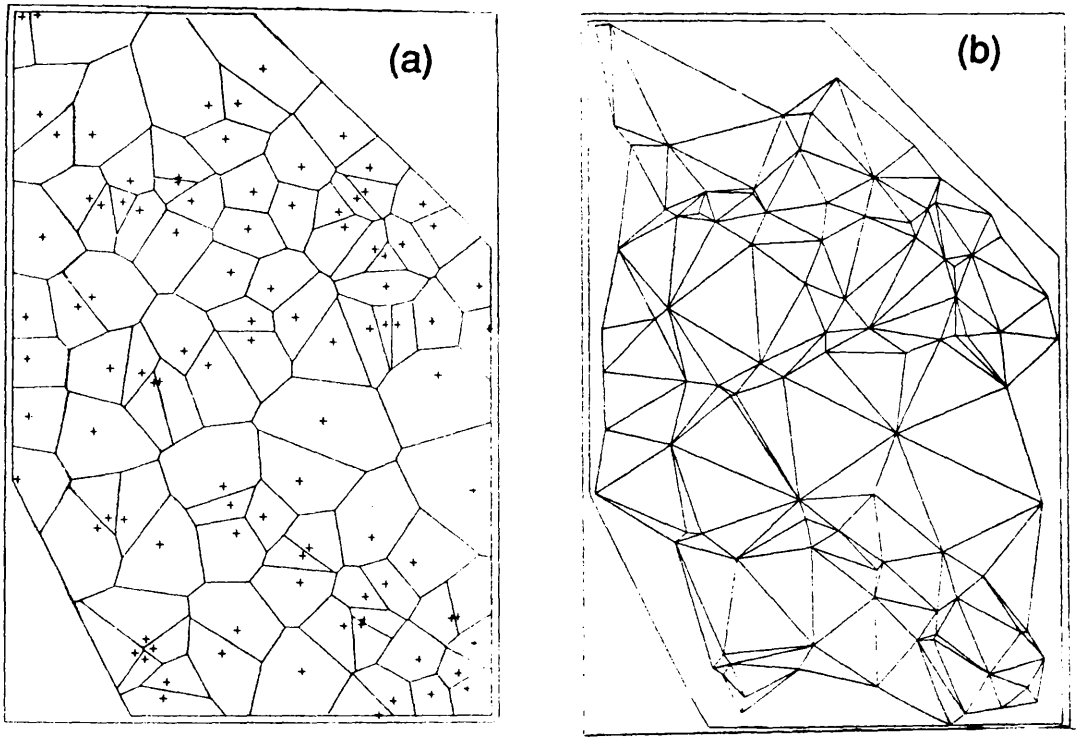


Fig.7.5 The dual relationship between Thiessen diagram and Delaunay triangulation (Green & Sibson, 1978). (a). The Thiessen polygons; (b). The correspondingDelaunay triangles

So, from the discussions above, it can be found that, at least in principle, an approach combining point-based modelling with either grid-based or triangle-based networks can be employed in the modelling of a surface. In practice, this type of discontinuous surface is not acceptable for topographic mapping. However, for some other purposes, such as planning, it could be a very efficient one for very flat areas, especially when high accuracy is not required. In other disciplines, such as statistical modelling, this type of surface is still widely used and the resulting map is referred to as "proximal" map, as in the case of Fig.7.3(b) which was produced by the Harvard SYMAP program. Fig.7.6 is a perspective view of the discontinuous surfaces constructed from irregularly located data produced by the Harvard SYMVU program (Ziegenfus, 1981).

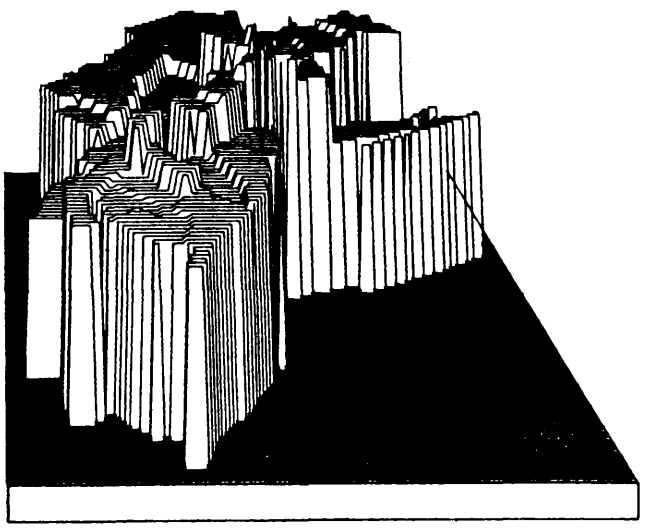


Fig.7.6 Discontinuous surfaces constructed from irregular data (Ziegenfus, 1981)

7.4.2 Continuous surfaces

By contrast, a **continuous surface** is one based on the idea that each data point represents a sample of a single-valued continuous surface. The first derivative of the surface can be either continuous or discontinuous. However, in this thesis, this definition will be used that a continuous surface is one which exhibits discontinuity in the first derivative. Furthermore, any surface which is continuous in both the first and in any higher derivatives will be referred to instead as a smooth surface.

Therefore, a continuous DTM surface will consist normally of a series of local surfaces (or patches) which are linked together to form a

continuous surface over the whole of the terrain area being modelled. The boundary between two adjacent sub-surfaces (or patches) may not smooth, i.e. it is not continuous in the first and higher derivatives. A series of contiguous linear facets is an example of such a surface. These have been employed extensively in practical DTM operation. Either the triangle-based modelling approach or grid-based modelling approach can be employed. Fig.7.7(a) & (b) show examples of such surfaces. Fig.7.8(a) is a cross section view of a boundary between two adjacent sub-surface, where a discontinuity occurs in the first derivative (Fig.7.8(b)).

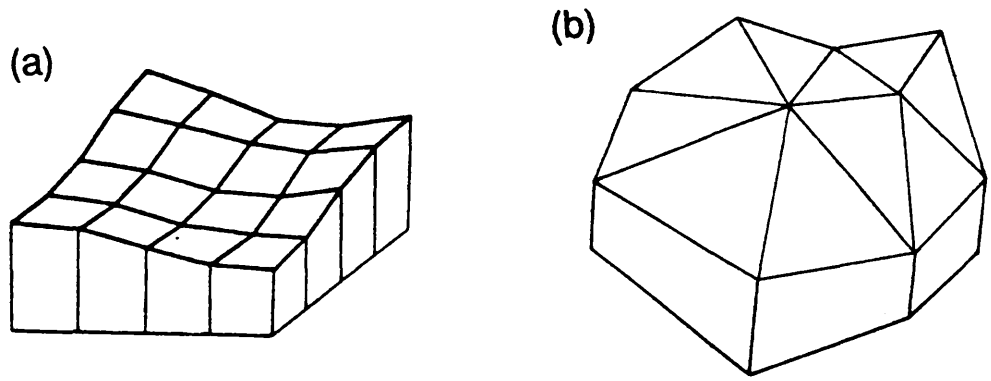


Fig.7.7 Continuous surfaces comprising a series of contiguous planar surfaces, based on (a) square grid cells (Ebner, 1980); (b) triangles.

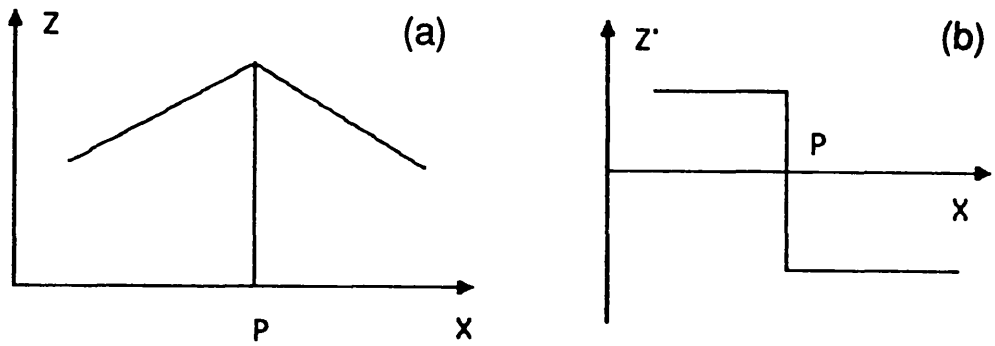


Fig.7.8 Discontinuity in the first derivatives at boundary point P
(a). Cross section view of a boundary between two adjacent facets;
(b). Discontinuity in the first derivative at the boundary.

The lack of continuity in the first derivative is, for many users of DTMs, something rather undesirable either in terms of the modelling itself or in terms of the final graphic output. However, the fact is also worth noting that the lack of continuity in the first derivative resulting in a distinct boundary between adjacent patches, grid cells or triangles is a feature which may not be undesirable in some cases. Indeed it may be deliberately sought after or introduced into the modelling process. This is particularly the case with data located along linear features such as rivers, break lines, faults, etc. acquired

via selective sampling or composite sampling. In such cases, the lack of continuity in the first derivative is indeed desirable so that interpolated contours change direction abruptly along such lines. Fig.7.9(a) is an example produced by SCOP.

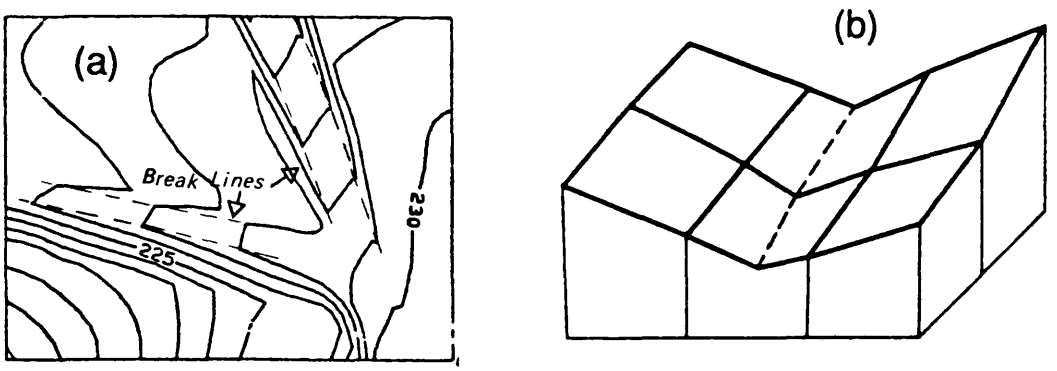


Fig.7.9 Continuous surface incorporating linear features
(a). Contours crossing break lines (SCOP);
(b). Linear feature in a grid-based surface (Ebner, 1980)

Even in a sophisticated grid-based DTM package such as HIFI, which incorporates smoothing and continuity between patches, provision is made to sub-divide an individual patch into two linear facets along the break line which acts as their common boundary allowing a continuous surface but without continuity in the first derivative (Fig.7.9b).

Furthermore, it can be found in the literature (Peucker, 1972), that, in many cases, a continuous surface comprising a series of contiguous linear facets is the least misleading one although it may look not so convincing or attractive from the visual point of view.

7.4.3 Smooth surfaces

Smooth surfaces refer to those surfaces which exhibit continuity in the first and higher derivatives. Usually, they are implemented on a regional level or on a global scale. The generation of such a surface is based on the assumptions that

i). the source data always contains a certain level of random noise or errors in measurement so that the surface to be constructed from the source data does not necessarily pass through all data points; and

ii). the surface to be constructed should be smoother than (or at least as smooth as) the variation which is indicated by the source data.

For this condition to be achieved, normally, a certain level of data redundancy is used and least squares method is implemented using a multi-termed polynomial to model the surface (Fig.7.10(a)).

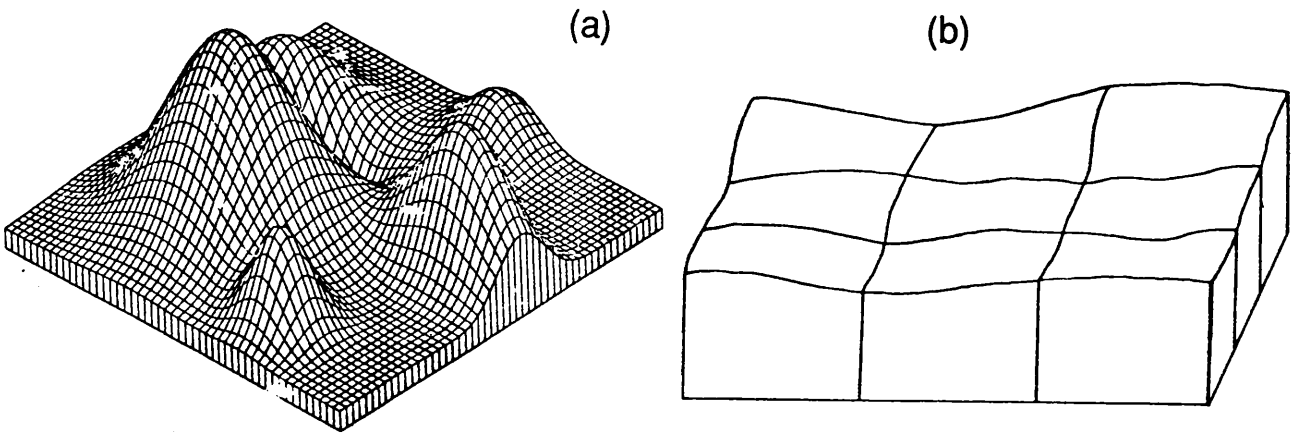


Fig.7.10 Examples of smooth surfaces
(a). A global (smooth) surface; (b). A smooth surface comprising a series of patches or regional surfaces

In the case of a single **global surface** based on a large data set, the whole of the surface is modelled by a single high-order polynomial. A huge amount of data may be involved, with an equation formed from each data point. However, the number of unknowns will still be relatively small, so that using a least squares solution, there will be a very substantial comuputational burden or overhead on the modelling operation when a very large amount of data is involved.

Actually, not only does this approach involve considerable computation in calculating the unknown parameters of the polynomial and in creating the model surface, but also, as stated before, the final result may often exhibit unexpected and unpredictable spikes or oscillations between data points. These are highly undesirable both in terms of the surface modelling process itself and in the fidelity of the final result in terms of the actual representation of the terrain surface delivered to the user in the form of contour plots or perspective views.

The result of these considerations is that the data sets are often divided into a series of **continuous patches.** The patches themselves may be regular in terms of shape and size, as in the case of square grid cells or equilateral triangles, or they may be irregular both in shape and size, as in the case of the randomly distributed points normally encountered in a triangulation procedure. Within each data patch, a lower-order polynomial can be used to model or reconstruct the surface again using the least squares method if redundant data is available.

While the use of the polynomial ensures a smooth surface within each patch, a break in continuity will almost certainly occur along the boundaries between patches since different terms will be dominant in the polynomials used in each of the individual patches.

The upshot of this is that continuity in the first and higher derivatives between adjacent patches will have to be built into the DTM system so that a smooth surface can be achieved between the patches without breaks or discontinuities along their boundaries. In other words, a so-called **seamless join** must result between patches. Needless to say, the successful implementation of such a requirement carries a very heavy computational overhead (Fig.7.10(b)).

7.4.4 Summary and discussion of surface continuity

Attempting to summarize the discussions made in this section, using the continuity of a surface as a basis for a classification of DTM surface, three types of terrain model surface can be distinguished, namely, discontinuous surfaces; continuous surfaces; and smooth surfaces. It has also been shown that, in most cases, discontinuous surfaces are not acceptable in the context of terrain modelling, although they are widely used in other disciplines such as statistical modelling.

By contrast to discontinuous surfaces, smooth surfaces always appeal to topographic scientists and to many practitioners or users of DTMs. However, care needs to be taken in the case of using high-order polynomials. Another point associated with smooth surfaces is that, in most cases, a heavy computational overhead may be involved if this approach is adopted.

It has also been discussed that, in many cases, continuous surfaces comprising a series of linear facets are the least misleading surface representation although they do not look so nice from the visual point of view. Another very important advantage of this type of surface consists in the flexibility of incorporating geomorphological information such as break lines, form lines, faults, etc.

As mentioned in the introductory chapter, it has also been found that the final accuracy of a DTM surface is most strongly affected by the attributes of the source data and by the characteristics of terrain but not usually by the type of model applied to represent the surface. Nevertheless, as will have become apparent from the discussion in this chapter, the selection of an appropriate method for surface modelling is a matter of some importance. As a result, some care has been taken to select appropriate models for use in the present project. In particular, the use of continuous surfaces comprising a series of linear facets obtained from both grid-based modelling and triangle-based

modelling will receive special attention in the assessment of DTM accuracy which will be discussed later in this thesis.

In some sense, these remarks could be the end of this chapter. Indeed, perhaps they should be. However, as stated before, just for the sake of completeness as a document, the methods used to form both triangular networks and gridded networks, which can then be used for terrain surface modelling by using the triangle-based approach and the grid-based approach respectively, will be briefly reviewed in the following two sections. In fact, another important reason for doing so is to give readers a more complete understanding about how the process of digital terrain surface modelling - from source data to network to surface model - is actually carried out in practice.

7.5 Triangular network formation

7.5.1 DTM networks: An introduction

As defined before, a **network** is a data structure implemented in a special pattern for surface modelling. It needs to be emphasized that a network concerns mostly the inter-relationship of the data points in the positional (planimetric) sense but not necessarily in the third dimension. This is the main **difference** between a network and the DTM surface which is constructed from the network and comprises a series of sub-surfaces which may or may not have continuity in the first derivative. The data structure for a regular grid is built-in (i.e. it is implicit) due to the special characteristics of the regular grid itself so that often this difference is not appreciated or shown clearly. By contrast, in the case of triangle-based modelling, this distinction is very clear - the planimetric inter-relationship needs to be sorted out to form a triangular network; then the third dimension can be added to the network to form a continuous surface comprising a series of contiguous triangular facets.

The triangular network pattern can be viewed as being the most basic one of all, since, as mentioned earlier, it can be applied to both regularly and irregularly located data. Also as already noted, a gridded network can be formed by interpolation from a triangular network and either a continuous or a smooth surface can also be constructed from the same network. Therefore, first of all, the formation of a triangular network will be discussed in the next sub-sections (Section 7.5.2 onwards) and a discussion about the formation of gridded network will be given later in Section 7.6.

The process of forming a triangular network is usually referred to as **triangulation.** A triangulation procedure can be applied either to

regularly distributed data such as gridded data to form a **regular triangular network** or to irregularly distributed data to form a TIN **(triangular irregular network)** which comprises a series of contiguous triangles of irregular size and shape.

7.5.2 Formation of a triangular network from regular data

If the source data has been acquired systematically in a regular pattern, then this is the simplest network of all to form. In the case of a square grid, simple sub-division using one or two diagonals produces a series of regular triangles. Fig.7.11a, b, & c show these three possible triangular patterns derived from a grid pattern. In the much less common case of the measuring pattern being based on regular triangles (Fig.4.1), then implicitly the network is already triangular.

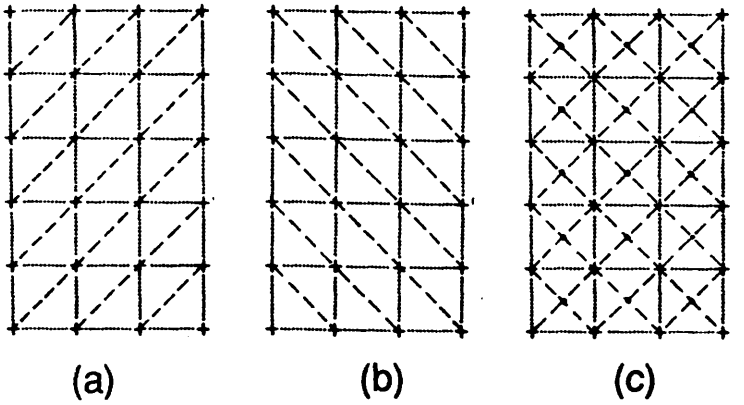


Fig.7.11 Triangular networks formed from a regular grid

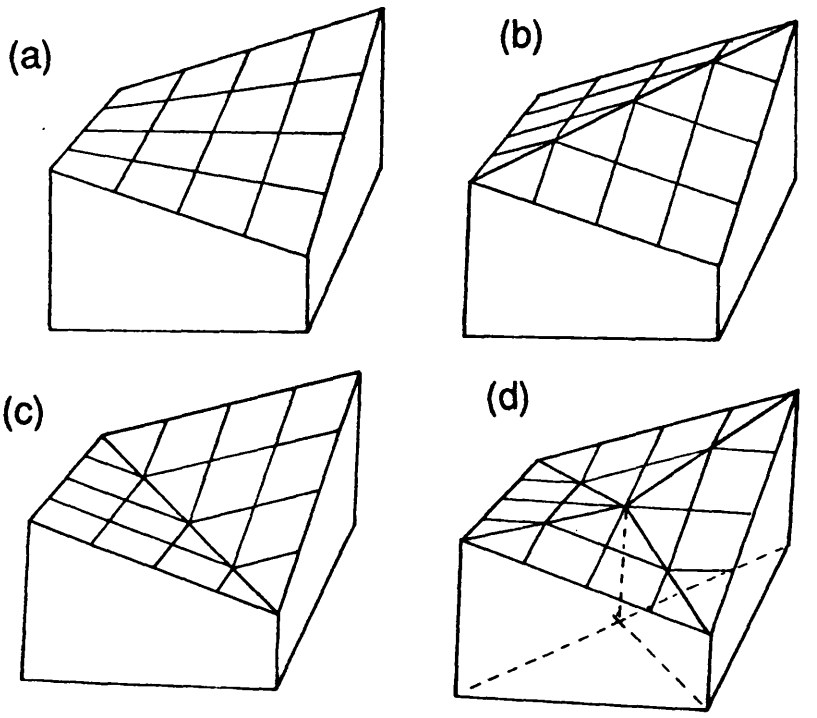


Fig.7.12 Possible types of linear facets constructed from a square grid cell

Of course, such an approach to form triangular networks from square grids is sometimes very arbitrary. Fig.7.12 shows such a situation clearly. Fig.7.12(a) shows a bilinear surface which can be constructed from a square grid cell. Fig.7.12(b) shows that such a grid cell can be split into two triangles by a single diagonal whose plan is shown in Fig.7.11(a). Similarly, Fig.7.12(c) shows the two triangles corresponding to those divided by the alternative single diagonal shown in Fig.7.11(b). Finally, those in Fig.7.12(d) correspond to the arrangement shown in Fig.7.11(c) with four centre point triangles formed by using both diagonals. It is apparent that the height value of a point interpolated from those different surfaces shown in Fig.7.12 (a) to (d) will all be quite different, although the same elevation values have been used at the grid nodes in each of the four examples given above.

Indeed, this is a problem with this type of arbitrary triangulation of square grids. Therefore, some care needs to be taken in practice. However, at present, most of DTM programs, even the commercial packages like CIP carry out such a triangulation process blindly (Fig.7.13).

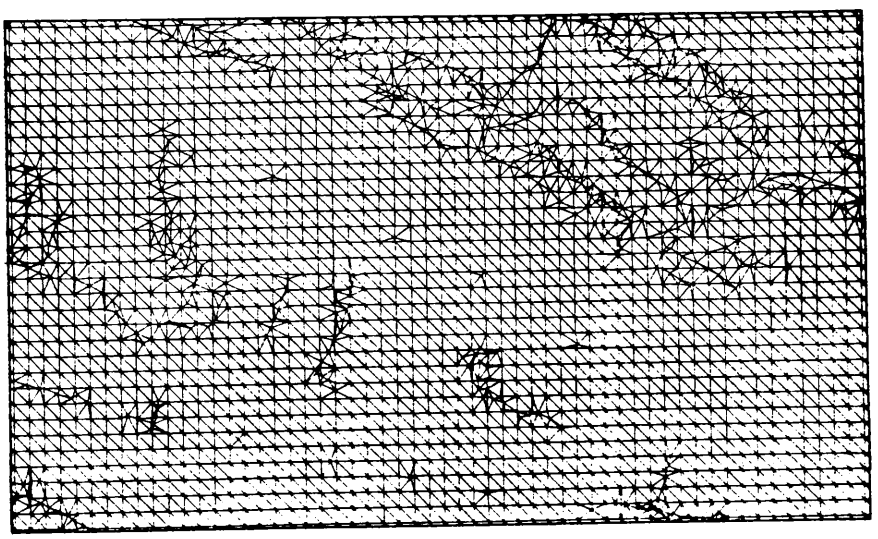


Fig.7.13 Triangulation from grids with linear features using hybrid approach (CIP)

7.5.3 Formation of a triangular network from irregular point data

If the source data is irregularly distributed, then special methods (or algorithms) need to be used. There are quite a few triangulation algorithms (methods) in common use. Their principles are briefly reviewed below.

(1). Radial sweep algorithm (RSA):

This algorithm was designed by Mirante and Weingarten in 1982. It

constructs a triangular network in two separate steps. The first step is to form a series of thin triangles radiating from a central point and the second step is to correct the shapes of these triangles until a more acceptably shaped series results to for the final network.

The point located nearest to the centroid of the data set is selected as the starting point for the network construction. Then the data points are sorted into an order which may be clockwise or anti-clockwise, according to their bearings around the starting point. After that, the radial line to each point is established and a thin triangle is formed using each pair of adjacent radial lines and the lines connecting the two data points at the ends of these two radial lines. The process of forming triangles is carried out systematically by sweeping around the starting point. The resulting triangles are very thin, but they are ordered, contiguous, and linked to their neighbours. Fig.7.14 shows the formation for these thin triangles.

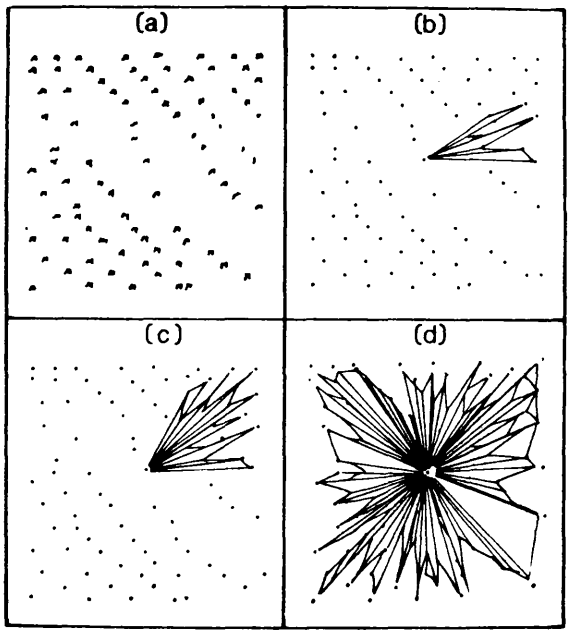


Fig.7.14 First step of forming triangles by RSA.

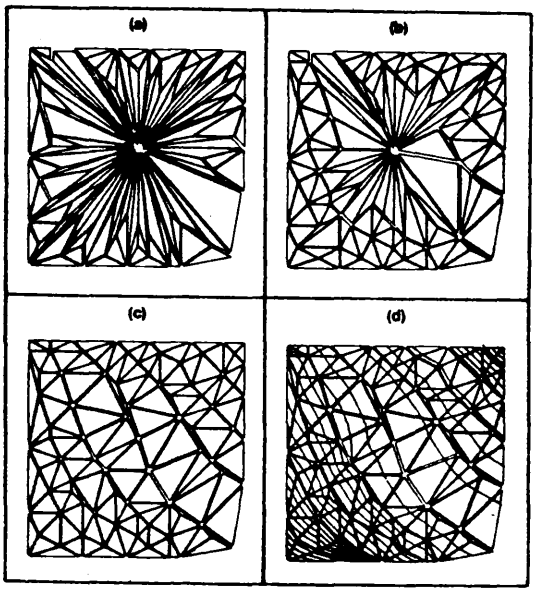


Fig.7.15 Concavity elimination and shape alteration for RSA (Mirante & Weingarten, 1982)

A supplementary step at this stage is also needed to form triangles to eliminate the concavities which exist around the perimeter of the data set. The next step is to alter the shapes of the triangles, i.e. to make triangles as nearly equilateral as possible. Two adjacent thin triangles may be altered by swapping their diagonals to form two new triangles which are more nearly equilateral in shape. Usually, several iterations of this process are necessary to achieve an acceptable network of triangles. If needed, then contours can also be produced from the DTM surface constructed from such a network. These two steps

can be summarized in Fig.7.15.

(2). Minimum sum-distance or circumscribing circle radius algorithms:

As the names imply, the sum-distance refers to the sum of the distances from a new point under inspection to the two end points of a base line. The corresponding algorithms are based on the criterion that the new point which should be selected to construct a new triangle is the one which has the sum of its distances to the end points of the base as the smallest value. Consequently, these can be referred to as **minimum sum-distance algorithms.** Similarly, the circumscribing circle radius refers to the radius of the circle which circumscribes the triangle consisting of two end points of a base and the new points under consideration. The algorithm is based on the criterion that the new point which is selected should form a triangle in which its circumscribing circle radius is the smallest value. Thus it is referred to as **minimum circumscribing circle radius algorithm.**

A number of existing algorithms are based on one or the other of these two criteria. In fact, they are very similar one to another.

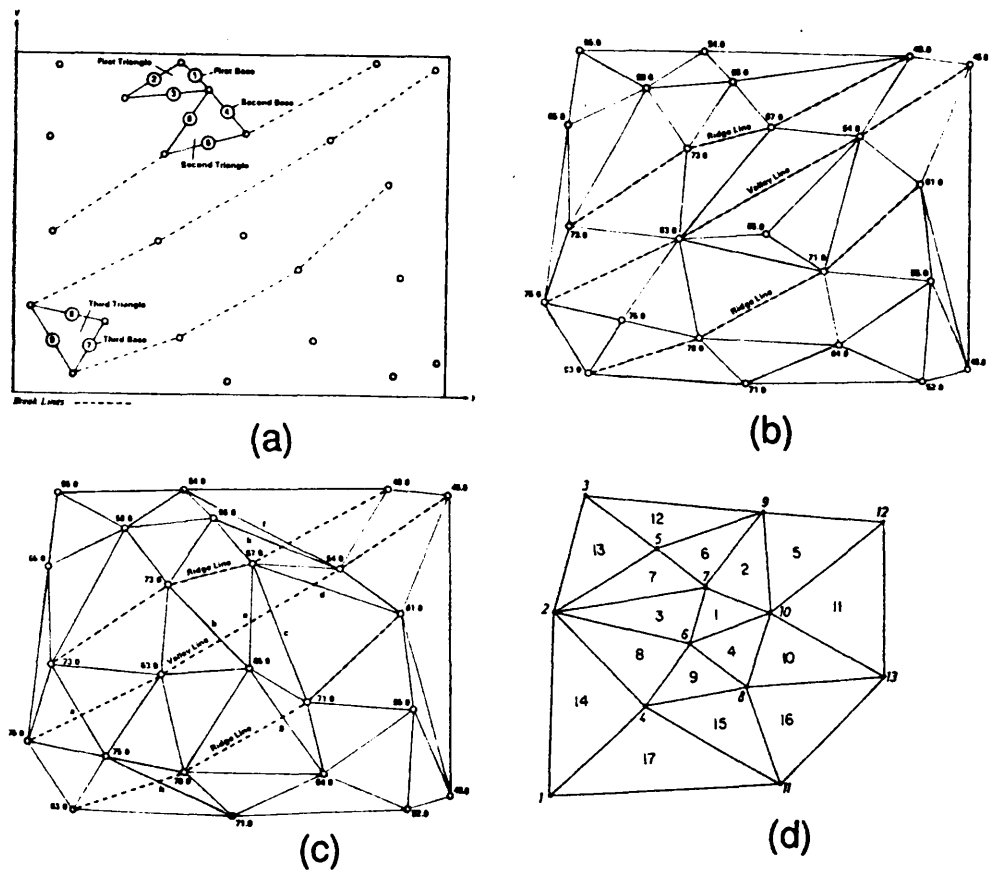


Fig.7.16 Triangulation with Yoeli's method
(a). The first three triangles; (b). Break lines provided;
(c). No break lines provided; (d). Layout of points and triangles

Yoeli (1977) developed a procedure, RANDCON (RANDom point CONtours), based on the **minimum sum-distance criterion.** His line of thought is as follows. First of all, the shortest of all possible lines between any two data points in the whole data set is searched for (Fig.7.16(a)). This line is then used as the base for the first triangle. The neighbouring points are then searched for and examined and, among these, the new point is selected as the third vertex of this triangle if the sum of its distances to the end points of the base is the smallest value among the sum-distances of all the neighbouring points which have been tested. After this, the second shortest line contained in the whole data set is searched for to serve as the base for the second triangle. This may not be contiguous with the first triangle (see Fig.7.16(a)). The process of forming other triangles then continues in a very similar way. However, any line which has already been used as a component of any previous triangle should not be selected as the base of any new triangle. Before any side of a new triangle is decided upon, the program checks that it does not intersect any already stored line or break line. If such an intersection occurs, the side is abandoned and an alternative searched for (Fig.7.16(b), (c), and (d)).

Elfick (1979) described a procedure to form triangles for the interpolation and plotting of the contours from stadia tacheometric survey data based on the **minimum circumscribing circle radius** criterion. The procedure is as follows: Firstly, the data points are sorted in ascending values of X; after this, the triangulation process can then be started. As Elfick (1979) has described, "The point nearest the centre is used as a starting point. It is then joined to the nearest point to it (to form the first base); and then the most suitable point to the right looking from the first point towards the second point is selected to form the third point of the first triangle. The most suitable point is that which forms a triangle with a circumscribing circle of minimum radius". The triangulation is then expanded from the three sides of the first triangle.

Other procedures can also be developed based on these criteria. For example, instead of starting from the point nearest the centre or the shortest line, the triangulation process can also start from the boundary of the area concerned using the concept of an advancing front (Elfick, 1979). The first front is just the boundary lines enclosing the data set. A procedure to form a triangular network based on this idea has already been developed in the Department of Photogrammetry and Geology, South-western Jiaotong University (in China), where the author used to work.

McLain (1976) has described an algorithm based on a criterion similar to that of the minimum circumscribing circle radius. However, instead of using the radius itself, the distance from the centre of the circum-

scribing circle to the base line is used in his algorithm.

(3). Delaunay triangulation:

The Delaunay triangulation method is the most popular one because of its consistency in terms of the resulting triangles which are produced. Regardless of the starting point of the triangulation process, the same set of triangles will be produced. As has been mentioned before, the Delaunay triangulation is associated with the Thiessen diagram. Delaunay (1934) first recognised that at the time when a Thiessen polygon is determined, a network of triangles can also be connected together, thus forming a triangulation. Thus a triangular network can be formed in the process of finding the neighbouring points to form Thiessen polygons.

In the original sense, the criterion of this method should be that the perpendicular bisectors of all sides connecting a starting point and its neighbours, which are found to form **Delaunay triangles**, should form a **Thiessen polygon** or reach to the boundary. However, it has also been pointed out in Section 7.4.1 that, nowadays, the inverse procedure is used in which the Thiessen polygons are defined from the Delaunay triangles. Therefore, forming a Thiessen polygon is not the criterion used in practice.

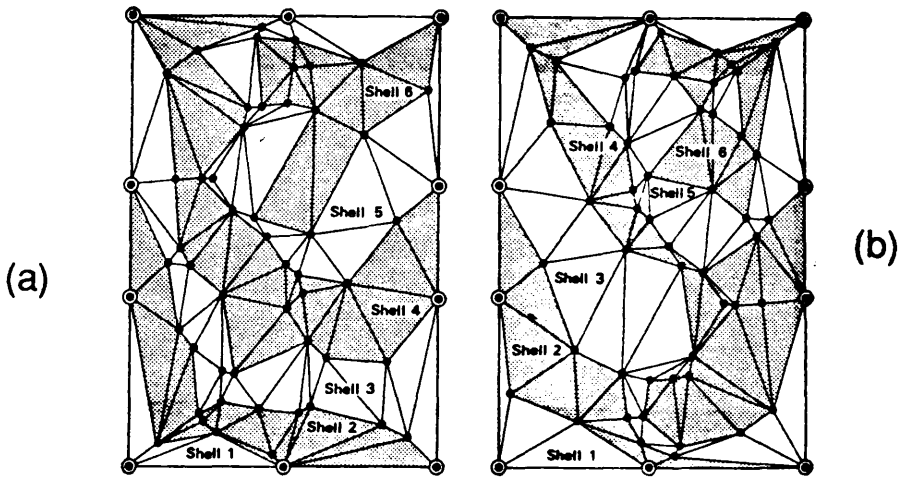


Fig.7.17 Results of Delaunay triangulation starting from two different imaginary points. (b) is rotated through 180^{o} against (a). (McCullagh and Ross, 1980).

There are possibly many procedures to form Delaunay triangles. The following is the procedure described by McCullagh and Ross (1980). This Delaunay triangulation procedure employs a set of imaginary points which serve to define the boundary box of the area to be triangulated. The triangulation process starts from two of these imaginary boundary points, although it may possibly begin using any arbitrary pair of points which are neighbours. In the latter case, one arbitrary point is

used as the starting point (also called the rotation point) of the triangulation procedure and another (the nearest neighbour) is called the known point. These two points form the initial base. Then a new point which is located to the right (i.e. clockwise looking from the known point to the rotation point) of the base and has the largest angle subtended from the base from all possible choices around the starting point is selected to act as the vertex of a new Delaunay triangle. Fig.7.17 shows the results of the Delaunay trianglulation starting from two different imaginary points, but with identical pattern.

Of course, these possible choices can be confined to a certain level by first defining a search area around the starting and rotation points using a circle with the base used as diameter. If this does not succeed in finding a point, then the search circle is expanded using the base line as a chord and with progressively larger circles until the appropriate neighbours are found (Fig.7.18).In this way, the most likely neighbours are first picked up and are then tested to find the one with the largest angle.

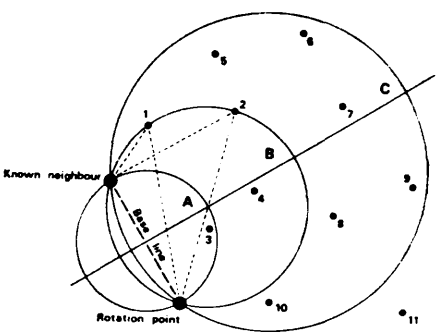


Fig.7.18 Search for nearest neighbour (McCullagh, 1983)

7.5.4 Formation of a triangular network from contour data

General purpose triangulation algorithms of the types discussed above consider every data point independently. If these were applied to digitized contour data, then the special configuration of contour data would not be taken into consideration. This would result in very poorly configured triangles and in some quite anomalous situations. These might include those where the three vertices of a triangle are taken from the same contour and that where a side of a triangle goes across a contour (Fig.7.19). For these reasons, special care has to be taken in carrying out the triangulation process with digital contour data. The implementation of a triangulation procedure which incorporates procedures needed to accommodate the special characteristics of digital contour data can be referred to as **contour-specific triangulation**

procedure. Actually, the main criterion which is used with such procedures is to treat every contour line as a break line or form line across which no triangle may be formed. Some procedures have also been developed for contour-specific triangulation such as those described by Gannapathy and Dennehy (1982) and Christensen (1987), etc. Fig.7.20 shows an example of a triangular network produced by Delaunay triangulation with special consideration for contours.

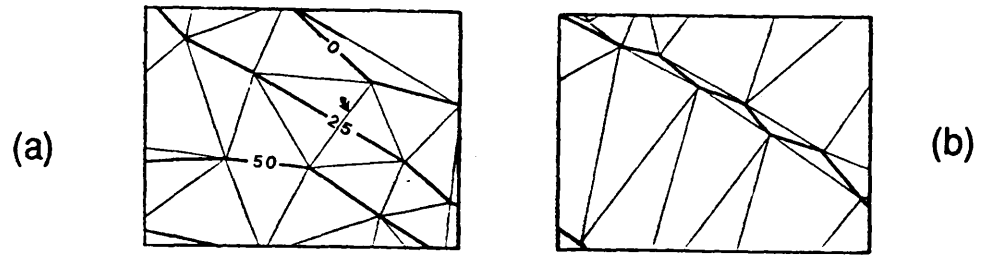


Fig.7.19 Problems with blind triangulation from contour data (Christensen, 1987). (a) Crossing a contour; (b). Three vertices of a triangle taken from the same contour line

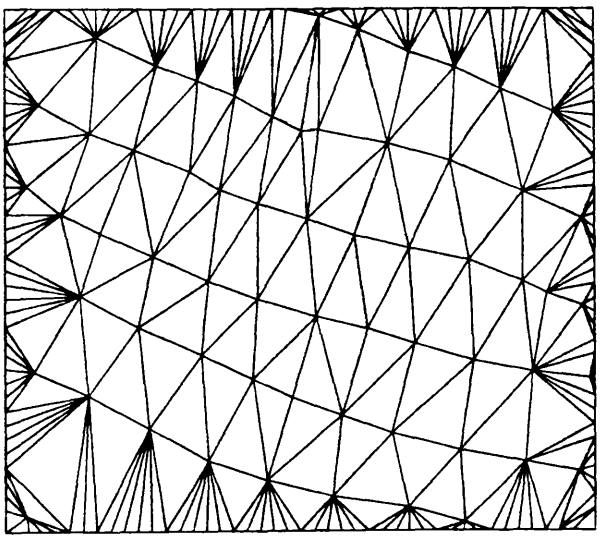


Fig.7.20 Triangular network formed by treating contours as break lines

7.5.5 Formation of a triangular network from composite data

Up to this stage, the formation of a triangular network from different data patterns - square grids, any irregular pattern as well as contour data has been considered. However, there is still one data pattern left to be mentioned here - that of composite data.

As has been discussed in Chapter 4, composite data could be a combination of string data (form lines, rivers and break lines) and gridded data either from regular grid sampling or from progressive sampling. A

typical example of the formation of a triangulation network from such a data set has already been shown in Fig.7.13 which was produced by CIP. That is, the individual grid cells have been split into regular triangles. However, if any linear feature passes through a square grid cell, then instead of splitting the cell using diagonals, the string of data points are also taken into account in the triangulation process to produce irregularly-shaped triangles within the grid cells.

7.6 Gridded network formation

7.6.1 Introduction

As has been discussed previously, grid-based surface modelling is the other main approach for constructing a DTM surface. However, rather obviously, this approach is only suitable for gridded data sets. Therefore, it is desirable to have a brief discussion about the formation of a gridded network from the different data patterns which are obtained using different sampling strategies.

It is obvious that, if the regular grid sampling method was employed on the basis of a square grid, then the resulting data is already in a suitable (grid) form. Therefore, no special process is needed to form the grid-based network.

However, if other sampling strategies such as selective sampling, profiling or contouring, etc. have been used during data acquisition, then comes the problem of how to form the required gridded network. In DTM literature, the process of forming a gridded network from any other non-grid data pattern is referred to as **random-to-grid interpolation.**

As discussed in Section 7.2 of this chapter, three methods were distinguished by Petrie and Kennie (1986) for random-to-grid interpolation, namely pointwise methods; patchwise methods; and global methods. However, basically only local information is of most interest in the process of random-to-grid interpolation, since, in this particular process, the concern is only with the adequate representation of the local elevation height by the surrounding grid nodes and not with the relief of the whole area to be modelled. Therefore, pointwise and patchwise interpolation methods are those which are most important for this particular purpose.

7.6.2 General methods

Normally, the use of **patchwise methods** involves the division of an area into a series of patches, preferably each of equal size and identical shape, and then constructing a local surface for each patch from all

the data points located within the patch by employing a relatively low-order polynomial function and finally assigning height values to each grid node lying within the patch in much the same way as is used in global methods. The interpolation of regular grid nodes from a triangular network can also be considered as a patchwise method where the individual patch is a triangle or a group of a few triangles.

Pointwise methods involve constructing a local surface for each point (grid node) to be interpolated and then assigning the height value of the local surface at the grid node position to this node. The local surface is constructed from the neighbours of the point to be interpolated. The local surface could be a summation of individual surfaces as discussed in Section 7.4.2 or a low-order polynomial surface fitting through the neighbouring data points, the centre of which is the point to be interpolated.

The **selection of the neighbours** may be specified in terms of the number of nearest points or those lying within a specified distance of a point to the points to be interpolated or some combination of both of these criteria. Almost always, a weighting function is used with most weight being given to the nearest neighbours. Thus, the weighting function could be a simple distance function (which is used in most cases) or a function determined by the distances between the neighbours (including itself) and the information about the height variation of this data set, e.g. using the variograms described in Chapter 3 in the case of Kriging method.

In the case of **"summation of surfaces"** method, these individual surfaces are usually level planes (i.e. each is a horizontal linear surface). From each of the measured randomly located neighbours, a horizontal linear surface is constructed. An average of all these level surfaces constructed from the neighbours may then be taken to give the elevation value at the node - often weighted according to the distance of each measured position from the node, leading to the name of the so-called **distance weighted average (DWA)** (Fig.7.21(a)).

An extension of this method forms **tilted linear (planar) surfaces** at each randomly located measured point, the tilt of each surface being determined by estimating the slopes in both the grid X and Y directions at this point - which could be derived by fitting a trend surface through the neighbouring points. Each separate linear surface is then projected on to the vertical line through the node and an average elevation value calculated for the node. Clearly this value will be different to that calculated using the series of horizontal surfaces. This method is called the **projected distance weighted average (PDWA)** value of the elevation at the grid node (Fig.7.21(b)).

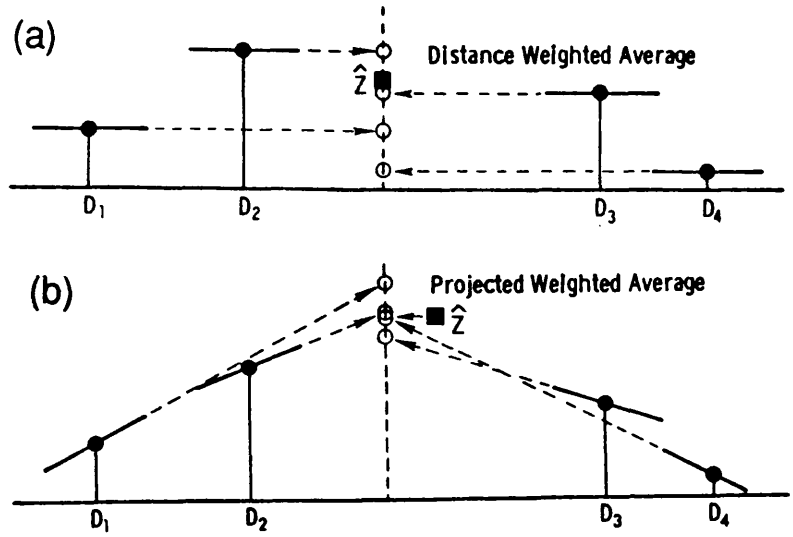


Fig.7.21 Distance weighted average for pointwise interpolation
(a) Cross section view of simple distance weighted average;
(b). Cross section view of the projected distance weighted average

In the case of using a low-order polynomial surface, the interpolation can be considered to take place in such a way that a local surface goes from one node to another with both its shape and orientation changed. Thus this method of random-to-grid interpolation is also referred to as the **"moving surface"** method.

Pointwise methods have also been implemented in other ways. In this context, either some functional models such as polynomial functions and sinusoidal functions as discussed before, or stochastic models such as those based on similarity and random functions (lFrederiksen et al, 1985), or a combination of these two such as ARIMA (AutoRegressive Integrated Moving Average) used in a Time series analysis (Lindberg, 1986) have all been employed. However, the simpler methods outlined in the previous paragraph are those which are most commonly used in practice.

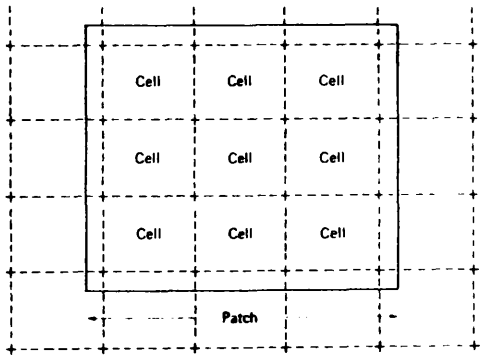


Fig.7.22 Nine cells as a patch for DTM surface construction

One point which needs to be emphasized here is that those gridded points obtained through random-to-grid interpolation are normally used

for grid-based surface modelling. Thus once these values have been obtained, they are usually followed by the construction of a continuous surface or smooth surface using the grid-based approach. In the latter case, more than one grid cell, normally nine grid cells (Fig.7.22), are will be used and a low-order polynomial function is fitted through these grid nodes to provide the required continuity in the first or higher derivatives or smoothing and to remove any ambiguity in the subsequent contouring of the data.

7.6.3 Grid network formation from contours

Gridded networks can also be formed from contour data. The process of carrying out the necessary interpolation to form the grid is referred to as **contour-specific interpolation** in this thesis.

In general, two different methods have been in use. One of them is the **contour-specific interpolation along certain pre-specified axes** (CIPA). The number of axes used may be one, two or four. The intersecting points formed by these axes and two adjacent contour lines are used as neighbouring points to be used for the interpolation. Then a pointwise interpolation is carried out by employing a distance-weighted function as before. Many papers about this method have been published such as those by Schults (1974) and Yoeli (1975), as cited by Clarke et al (1982). As shown in Fig.7.23, all of these points from 1 to 8 will be used as the reference points for interpolating the height value of point P, in this case.

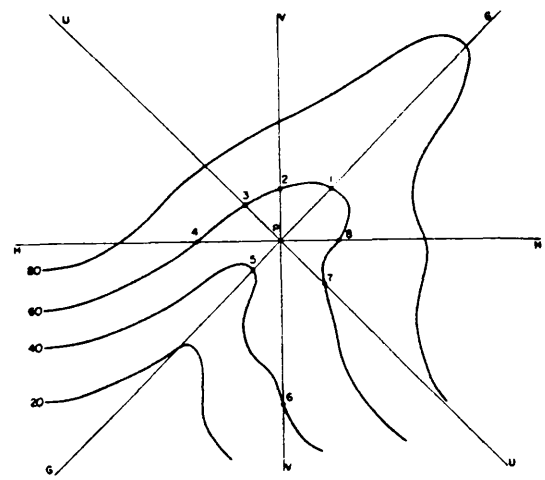


Fig.7.23 Grid point interpolation from contour data
(Leberl and Olson, 1982)

Another method is called **interpolation along the line of the steepest slope** passing through the point to be interpolated (IASS). Unlike CIPA, but like the procedure of manual interpolation, two points on the

adjacent contour lines along the **steepest slope** line are searched for and used for interpolation. Then the elevation value desired for the grid node position is interpolated linearly from these two points. As described by Leberl and Olson (1982), all those points (1 to 8) intersected by the pre-defined axes on two adjacent contours (at heights of 60m and 80m in this diagram) are used to determine which of these directions has the steepest slope. In this case, points 1 and 5 (Fig.7.23) will be used for linear interpolation of the height value of point P. Non-linear interpolation using cubic polynomial function has also be implemented by Clarke et al (1982). In this case, as shown in Fig.7.24, four points on the four contours (two next to each direction of either up or down) along the steepest direction are used. The method is called **cubic interpolation along steepest slope** (CISS).

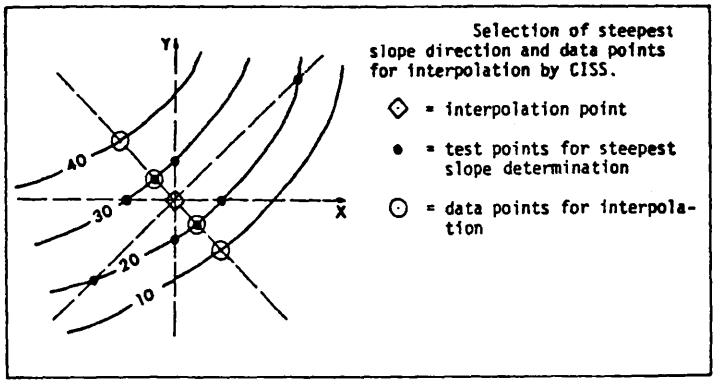


Fig.7.24 Points used for CISS method (Clarke et al, 1982).

Actually, the problem associated with all contour-specific interpolation is how to find the reasonable points for interpolation. In this respect, Inaba et al (1988) tried to use the aspect information of contours.

7.7 Concluding remarks

In this chapter, some basic concepts in surface reconstruction have been introduced and defined; the approaches which could be used in surface modelling have also been discussed; the basic methods used in these approaches have also been pointed out; different types of surface which it is possible to construct from a data set have been sketched; after which, the methods for forming both a gridded network and a triangular network have also been reviewed.

Up to this stage, the first part of this thesis - the theoretical background - has been described and discussed. From the next chapter on, attention is turned to the author's own main investigations which are more practically orientated. The discussion will start considering the quality of source data.

Chapter 8

Techniques for the improvement of the quality of DTM source data

Chapter Eight

Techniques for the Improvement of the Quality of DTM Source Data

8.1 Introduction

As one can imagine, the quality of the DTM source data will greatly affect the fidelity (or accuracy) of the final DTM surface which is constructed from the source data and the products derived from the DTM. Therefore, before starting any discussion about the accuracy of final DTMs, it is pertinent to discuss first of all the matter of the quality of DTM source data.

The **quality** of the DTM source data can be judged by using its three attributes (i.e. accuracy, density and distribution) as criteria. In the first place, it is apparent that the quality of a set of data can be considered as being **poor** if the data points are not well distributed, e.g. in a situation with very few scattered points in areas of rough and steep terrain but with a high density of points distributed in very tight clusters in relatively smooth and flat areas. However, these first two factors - **density** and **distribution** - are related to sampling, which has been discussed in Chapter 4, and the problems related to both of them can be solved somehow by employing an appropriate sampling strategy. Thus they will not be discussed further here.

Another important factor concerning the quality of DTM source data is its inherent **accuracy**, which has been discussed in Chapters 5 and 6. It is also obvious that, the lower the accuracy, the poorer the data quality. Accuracy is related primarily to **measurement**. After a set of data points have been measured, an accuracy figure can be obtained or estimated. Here it needs to be emphasized that the accuracy figure obtained (or achieved) for any measured data set is the overall result of different types of errors. In fact, the aim of this chapter is to devise algorithms or procedures to eliminate or to reduce the effects of some of these errors to achieve an improvement in the quality of the final DTM and thus in its products.

8.2 Errors in DTM source data

It is well-known that any measured data set is subject to various types of **errors** regardless of the method of measurement. In this respect, DTM source data is no exception. These errors affect the quality of the DTM source data. Thus, the problem of dealing with these errors will arise.

Generally speaking, three types of errors can be distinguished, namely,

random errors (also referred to as random noise in image processing and as white noise in statistics), systematic errors and gross errors.

In classical error theory, the variability of a series of measurements of a single quantity is due to observational errors. Such errors do not follow any deterministic rules, thus leading to the concept of **random errors.** Rondom errors have been recognised as having a normal distribution. A statistical value such as RMSE (root mean square error) or SD (standard deviation) can be calculated from these measurements to represent the precision of the set of measurements.

In this chapter, an investigation into the effects of these random errors in source data on the quality of the DTM and its products will be carried out and the results obtained from an experimental test using a set of photogrammetrically measured data will be described.

In photogrammetric practice, **systematic errors** usually occur due to physical causes (e.g. they may be the result of temperature changes in the photographic materials or in the instrument itself); or they may arise from the lack of adequate adjustment of the instrumentation before use. Alternatively, they may be the result of the human observer's limitations, e.g. in stereo-acuity or through carelessness such as failing to execute a correct absolute orientation. Systematic errors may be constant or counteracting. In reality, most practitioners in the area of terrain data acquisition are well aware of systematic errors and usually strive to minimize their presence. Therefore, no attempt has been made to deal with this type of error in the present study.

Gross errors are, in fact, mistakes. During measurement, they should occur with a small probability. Gross errors happen when, for example, the operator makes or records a wrong reading on the correct point, or observes the wrong point through misidentification; or the measuring instrument is not in proper working order when an automatic recorder is used. Gross errors often occur in automatic image correlation due to mismatches of the images.

From a statistical point of view, gross errors are specific observations that cannot be considered as belonging to the same population (or sampling space) as the other observations. Therefore, they should not be used together with the other observations from the population. Consequently, measurements should be planned and observational procedures designed in such a way as to allow for the detection of gross errors so that they can be rejected and removed from the set of observations (Mikhail, 1976).

Therefore, in this chapter, some algorithms will be described later for

detecting those gross errors which occur in DTM source data. Experimental results produced from the use of these algorithms on some sets of photogrammetric data will also be presented.

8.3 The improvement of DTM quality with filtering processing

As mentioned above, some algorithms or procedures will be developed to deal with the various errors existing in DTM source data. In this section (8.3), a discussion of the effects of the random errors and other high frequency phenomena present in source data on the quality of the DTM derived from it (the source data) will be given. Then a procedure designed to reduce these effects will be described and some results given.

8.3.1 Random errors, data components, random noise and filtering

In this context, any high frequency phenomena (i.e. small-scale variations in the measured data) may be regarded as **random noise.** So random error can be regarded as a type of random noise. Alternatively, the view can be taken that random noise includes both random errors and also other high frequency phenomena present in a data set.

In order to understand these other high frequency phenomena present in a data set, it is desirable to have a look at the data components. Any spatial data set can be viewed as consisting of three **components,** namely, (i) regional variations; (ii) local variations; and (iii) random noise. In the context of digital terrain modelling, the first component is of most interest since it defines the basic shape of the terrain surface. Interest in the second component varies with the scale of the desired DTM product. For example, at large scales, it is extremely important. However, if a small-scale contour map covering a large region is the desired product, then the second component may be regarded as random error since much less detailed information about the spatial variation of the terrain surface will be needed in this case. By contrast, it is definite that the third component is always a matter of concern since it may distort the image (appearance) of both the regional and local spatial variations in the terrain surface, but especially the latter. As a matter of fact, it is very difficult to define these data components clearly. Therefore, as discussed above, random noise, in this particular context, will be regarded as the high frequency component of the data set.

It is obvious that it is important to separate the main components of the data set which are of interest to the user from the remainder of the information present in the data set which is regarded as random noise. The technique used for this purpose is referred to as **filtering**

and the device or procedure used for filtering is referred to as a **filter.** Thus the process of applying a filter to a data set is referred to as **data filtering** in this thesis.

After these definitions, it is clear that a digital filter can be used to extract a particular component from a digital data set, thus ensuring that all other components are filtered out. If a digital filter separates the large-scale (low frequency) component from the remainder, this filter is called a **low-pass filter.** By contrast, if a digital filter separates the small-scale (high frequency) component from the remainder, then it is referred to as a **high-pass filter.** However, in this study, only the low-pass filter is of interest since it is the high frequency component that needs to be filtered out.

In this section, the effect of random noise on the quality of the DTM (and thus on its products) will be examined first of all; then a low-pass filter will be designed and used to filter out the high frequency component present in a digital data set to see if any improvement in data quality will result.

8.3.2 The effect of random noise on the quality of DTM source data

Before conducting a discussion about how to filter out the random noise and how much improvement in the quality of the DTM data will result after the application of a filtering process, it may be pertinent to quote some examples from existing literature showing how random noise affects the quality of the DTM and thus its products.

Ebisch (1984), discussed the effect of the round-off errors found in gridded DTM data on the quality of the contours which were derived from it, and he also demonstrated the effect of random noise in the DTM data on the contours produced from it. Ebisch first produced very smooth contours (Fig.8.1a) with 1.0m intervals from a conical surface which is represented by a grid of 51 by 51 points. Then he rounded off the grid heights to the nearest 0.1m (1dm) to produce another contour map (Fig.8.1b) to show the effect of round-off error. After that, he added random noise with a maximum amplitude of $\pm$0.165m to the grid heights and produced another map which resulted in zigzag and meandering contour lines (Fig.8.1c). This example shows very well the effects of random noise on the quality of DTM source data, and thus on the quality of the contours derived from this data.

Fig.8.2(a) is another example (see Section 8.3.4) showing the effect of random noise in DTM source data on the quality of the contours produced from it. It can be seen clearly that, from the visual point of view, these contours are not at all desirable or acceptable due to the effects of random noise, which include both the measuring error and

(perhaps) too detailed information about the terrain roughness present in the data set, resulting in high frequency local variations.

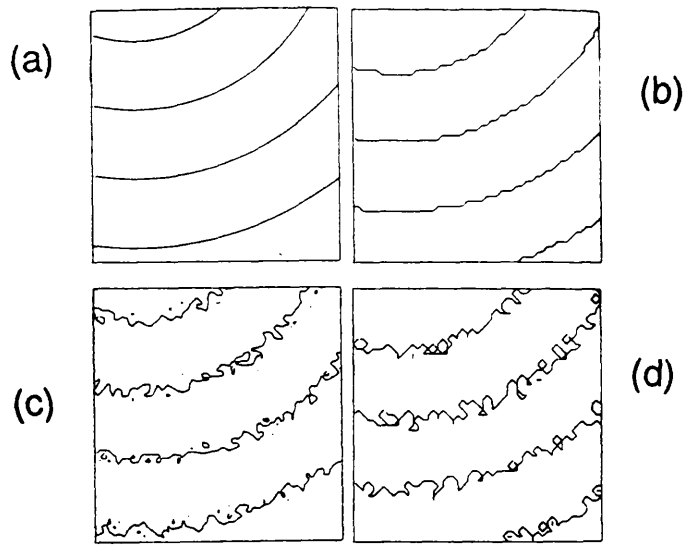


Fig.8.1 Effect of round-off errors and random noise on the contours produced from a DTM data set

(a). Contours produced from the original data set (a smooth surface);
(b). Contours produced from the data set after rounding off the data heights to the nearest dm;
(c). Contours produced from the data set with a random noise of magnitude $\pm$0.165m added.
(d). Contours produced from the data set with both random noise and round-off errors included.

8.3.3 Design of a low-pass filter based on a convolution operation

The effects of random noise in DTM data on the quality of the contours produced from the DTM data have been demonstrated above. In this section, a low-pass filter based on a convolution operation will be described which will be used later for data filtering.

(1). Definition of convolution

Convolution can take place either as a 1-D or a 2-D operation. However, the principles are the same in both cases. Therefore, for simplicity, the 1-D convolution is discussed here.

Suppose X(t) and f(t) are two functions, and the result of convolving X(t) with f(t) is Y(t). Then the value of Y(t) at position u, is defined as follows:

$$Y(u) = \int_{-\infty}^{+\infty} X(t)\, f(u-t)\, dt \qquad (8.1)$$

In the case of the DTM data filtering carried out in this study, X(t) is the input data in which random noise is present; f(t) can be thought of as a normalized weighting function; and Y(t) comprises the low frequency components present in the input data and can be thought of as the remaining part after filtering out the random noise. Actually, Y(t) is a smoothed function. Practically, it is not necessary to have the integration from minus infinity to positive infinity for Formula (8.1). In most cases, an integal which operates over a certain restricted length will do.

Certain functions such as a rectangular pulse, a triangular pulse or a Gaussian pulse can be used as the weighting function for this purpose. In this experiment, the Gaussian pulse has been used. The expression is as follows:

$$f(t) = Exp(-t^2/2Var) \qquad (8.2)$$

where Var is the variance of the Gaussian distribution.

(2). Discrete convolution operations

The definition of convolution given above is that which applies to continuous functions. However, in DTM practice, the source data is only available in a discrete form such as a regular grid. Therefore, only the **discrete convolution** operation is of interest here. The principle of such an operation using a symmetric function as weighting function will be described here since the weighting function used in this experiment - the Gaussian pulse - is symmetric. Its principle as applied in 1-D is explained below.

Table 8.1 Discrete convolution operation

X(t)	00	00	A1	A2	A3	A4	A5	A6	A7	00	00	Results	
Operations	x	+ x	+ x	+ x	+ x	+ x	+ x	+ x	+ x	+ x	+ x		
f(t)	W1	W2	W3	W4	W5							=	B1
		W1	W2	W3	W4	W5							B2
			W1	W2	W3	W4	W5						B3
				W1	W2	W3	W4	W5					B4
					W1	W2	W3	W4	W5				B5
						W1	W2	W3	W4	W5			B6
							W1	W2	W3	W4	W5		B7

Suppose, X(t) = (A1, A2, A3, A4, A5, A6 and A7);
f(t) = (W1, W2, W3, W4 and W5); and
Y(t) = (B1, B2, B3, B4, B5, B6 and B7);

then, the convolution operation is illustrated in Table 8.1. To explain how it works, the result for B4 can be taken as an example, i.e. B4 = W1xA2 + W2xA3 + W3xA4 + W4xA5 + W5xA6.

(3). Parameters for the filter

The size of the window and the weights selected for the various data points lying within the window have a large effect on the degree of smoothing achievable by the convolution operation. For example, if only one point is present within the window, then no smoothing effect will actually take place.

Also the smaller the differences in the weights given to the points lying within the window, the larger the smoothing effect it will have. For example, if the same weight is given to every point within the window, then the result is simply the arithmetic average. Table 8.2 lists some of the values for the Gaussian pulse expressed by Equation (8.2). From these values, a variety of weighting matrices may be constructed. Of course, the weight matrix can also be computed directly from Equation (8.2) using pre-defined parameters.

Table 8.2 Sample values of the Gaussian function

t	0.0xSD	0.5xSD	1.0xSD	1.5xSD	2.0xSD	3.0xSD
f(t)	1.0	0.8825	0.6065	0.3247	0.1353	0.0111

8.3.4 Experimental test

The source data used in this study was generated using a completely digital stereo-photogrammetric system (DSP) which has been designed and implemented by one of my research colleagues, Mr. Ali Azizi (1990). The digitised photos used in the DSP were formed from a pair of aerial photos taken at scale of about 1:18,000 using a scanning microdensitometer with a pixel size of 32um. The DSP system utilizes the image coordinates primary solution to form the stereo-model. The data was measured in a profiling mode with a four pixel interval between measured points. Thus the interval between any two data points is 128um at photo scale. Since this DSP is based on an image coordinate primary solution, the data points, which were arranged in grid form (with a 4 pixel interval) on the image plane, produced a data set only approximately in a grid form in this test area, with a grid interval of about

2.3m. The data generated by this system for this study is very dense. In an area of about 1cm by 1cm at photo scale, the elevations of 8,588 (113x76) points were measured (Azizi, 1990). This data set provides very detailed information about the terrain roughness.

The check points used for this study were measured from the same photos in hardcopy form using an analytical instrument (Azizi, 1990).

A filter based on the convolution operation described above was used for this test. Since the data is not in an exact grid, a 1-D convolution was carried out in each of the two grid directions rather than a single 2-D operation. Therefore, for every point, the average of the two corresponding values is used as the final result.

The window size comprises 5 points along each grid direction. The five weights for these five points were computed according to Equation (8.2) individually since the point intervals are varied. These values before normalization are approximately as follows:

$$f(t) = (0.1353, 0.6065, 1.0, 0.6065, 0.1353) \qquad (8.3)$$

In computing the value for each of these five weights corresponding to each of the five points lying within the window, the distance of the point to the central point of the window is used as the value of the variable t. Also the average value of the interval between each pair of data points (i.e. 2.3m) was used as the standard deviation of the Gaussian pulse - the weighting function.

Table 8.3 Accuracy improvement with random noise filtering

Parameters	Before Filtering	After Filtering
+ Max residual	+ 3.20 m	+ 2.67 m
- Max residual	- 3.29 m	- 2.76 m
Mean	0.12 m	- 0.02 m
SD	±1.11 m	±0.98 m
RMSE	±1.12 m	±0.98 m
No. of Check Pts	154	154

Where, Mean denotes the mean value of the DTM errors;
SD denotes the standard deviation of the DTM errors; and
RMSE is the root mean square error.

A triangulation-based DTM package (PANACEA) was used for this experiment - both for the interpolation of the DTM points and for production of contours from the DSP data. The interpolated height values were compared with the corresponding check heights and then the standard deviation (SD) and the root mean square error (RMSE) values were computed. The results for the data set both before and after filtering are given in Table 8.3, where it can be seen clearly that much improvement in accuracy (about 14% in terms of the RMSE value) has been achieved by using a filtering process on this particular data set.

Contour maps were also produced for comparison. The contours were threaded directly from the triangulated network. The contours produced from the original DSP data are shown in Fig.8.2(a). The corresponding contours from the data set after the filtering process are shown in Fig.8.2(b). It can be clearly seen that the small fluctuations in the contour shapes arising from the noise in the data have to a large extent been removed. Therefore, the presentation of the contours after the filtering processing is also much better from the visual point of view.

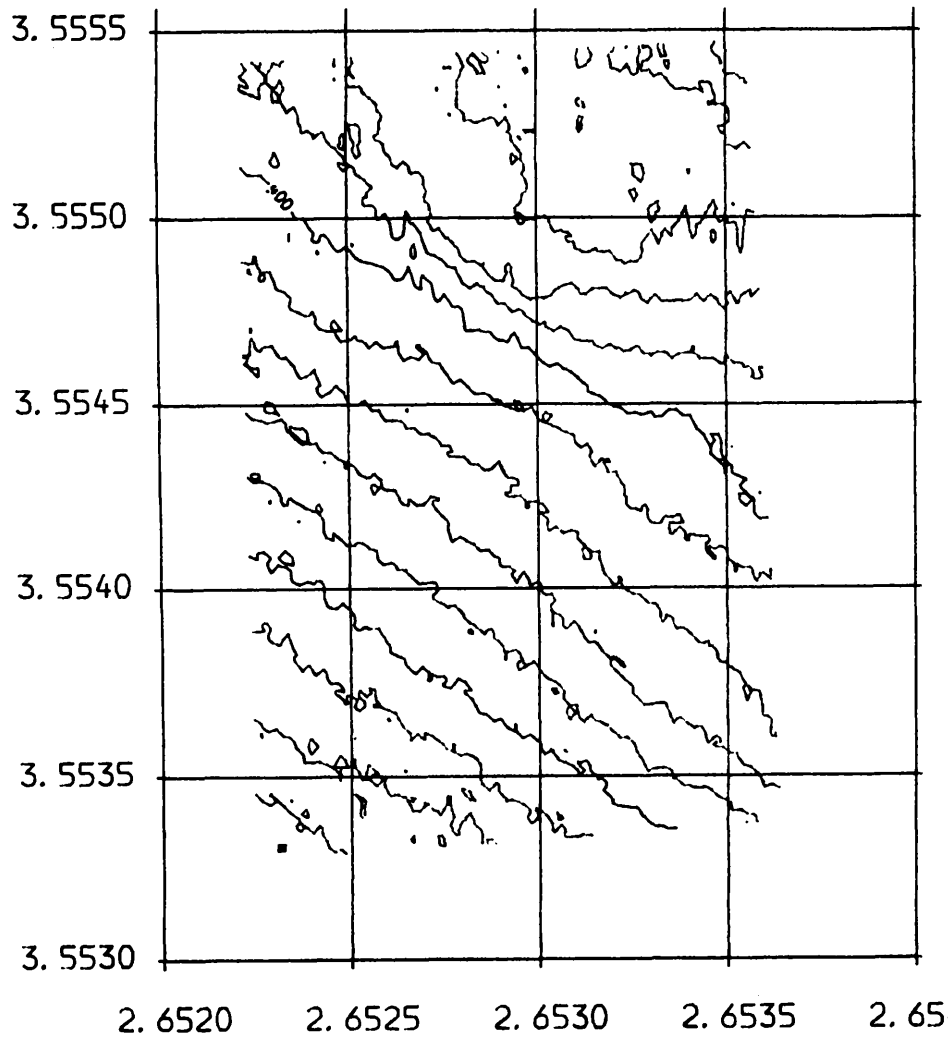
8.3.5 Discussion of data filtering regarding data quality

The data set used in this particular test is very dense. The average grid interval is 128um at photo scale which corresponds to about 2.3m in the terrain. Realistically, such a dense data set can only be obtained from devices equipped with automated or semi-automated techniques, e.g. using image matching techniques for height measurement based on automatic image correlation. In such a data set, the loss in the fidelity of representation of the terrain topography is not likely to be a serious problem. By contrast, the effects of the random errors involved in the measuring process and of any other random noise on the data quality is considerable at the local or detailed level.

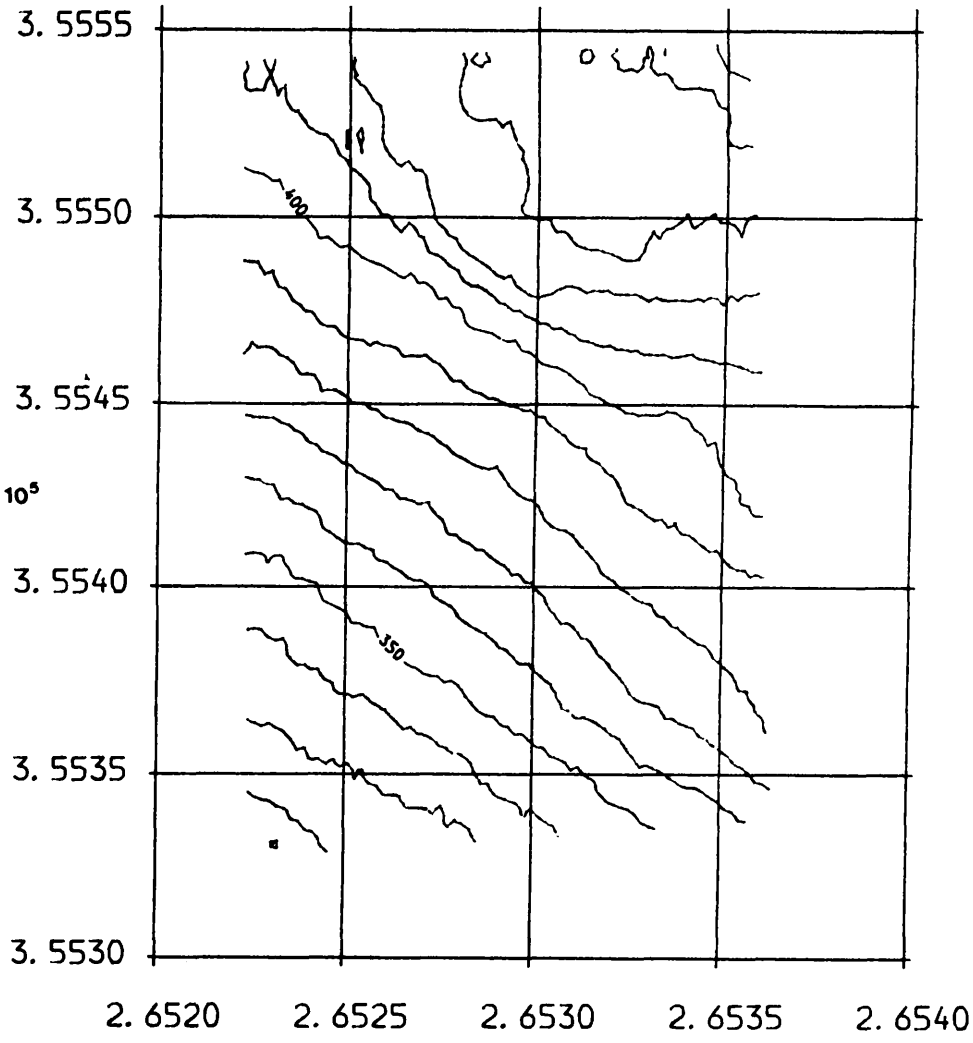
As quoted previously in Chapter 5, 70 to 90% of stereo-plotting machine operators are measuring with a precision (RMSE) within the range $\pm$10um to 20um (Schwarz, 1982). Such a precision of the parallax measurement will result in a precision (RMSE) of $\pm$0.3m to 0.6m for the measured data in this particular case. For this relatively smooth and homogeneous area, the random errors associated with such a precision of measurement will, of course, distort the spatial variation considerably.

By comparing the RMSE values resulting from the data set before and after filtering, it can be computed (according to the law of error propagation) that the amount of improvement is about $\pm$0.54m. This value is well within the empirically determined value range as given above.

From the discussions above, it is very clear that the availability of



(a)



(b)

Fig.8.2 Improvement of data quality using a low-pass filter
(a). Contours produced from the original data set;
(b). Contours produced from the smoothed data set

too detailed information about the roughness of the terrain morphology, coupled with the measuring errors likely to be encountered with image correlation techniques, can have a significant negative effect on DTM quality, and thus on the quality of derived DTM products such as contour maps. In the case of very dense data, a filter such as that based on a convolution operation which has been implemented in this study can be used to smooth the digital data set, and improve the quality of the digital data. Thus the quality of the products derived from this data can also be much improved.

A very important **question** arising from the results of this test is: when should a filtering process be applied to the digital data? This is, of course, a question which is very difficult to answer. Obviously, the magnitude of the random errors occuring during measurement need to be taken into consideration. According to the remarks made by Dr. Forstner (see Kubik and Roy, 1986), such a value might be considered as smaller than $\pm$0.05 per mil of H. Therefore, a **rough answer** to this question might be that, if the accuracy loss arising from data selection and reconstruction (topographic generalisation) is much larger than this value ($\pm$0.05‰ of H), then the filtering process should not be applied. By contrast, if the random noise does form an important part of the error budget, then a filtering process may be applied to improve data quality.

As can be seen from inspection of Fig.8.2(b), there are still some contours with segments which exhibit zigzag patterns. This may be due to larger errors occuring during measurement or they may be the results of other unnatural features associated with the terrain variations themselves. This problem will be considered in the next section.

8.3.6 Discussion of the computational effort required for data filtering

As stated in the introductory chapter (One), the cost and efficiency of processing are another two factors (besides accuracy) which are of great concern in the context of digital terrain modelling. Therefore, it is pertinent to have a brief discussion regarding the computational aspects of data filtering. Obviously, if the computational effort involved in data filtering places a large overhead and expense on the overall DTM operation, then the filtering process is unlikely to be used. Thus some quantitative values are needed in order to form a judgement on this particular matter.

(1). Relationship between cpu time and number of data points

The smoothing procedure described in Section 8.3.3 was implemented on the ICL 3980 mainframe computer belonging to the University. For this

particular example of 8,588 points, it takes 33,369ms cpu time. This is equivalent to 3.89ms per point, which is a relatively fast processing time.

As one can imagine, much of the computation time must have been spent on the **calculation of weights** for this particular procedure. But in practice, the weight function can be pre-defined if the data is gridded. Therefore, an attempt was made to see how much cpu time can be saved if the weights are pre-defined instead of being calculated. Therefore, another test was carried out, in which the weight matrix given by Equation (8.3) was used. As expected, a much smaller amount of cpu time - 16,645ms - was used. This is equivalent to 1.94ms per point. Even in the multi-user environment of the ICL 3980 mainframe machine, the results from this test came out almost instantly.

It seemed very reasonable to assume that the cpu time needed for a specific data set is **linearly** proportional to the number of points in the data set since the same amount of computation is involved for each point. To confirm this point, another smaller set of data comprising of 1,862 (49x38) points was also processed. Again, the weight matrix given by Equation (8.3) was used. The resulting cpu time was 3,506ms. It is equivalent to 1.89ms/point.

The results obtained from these two tests seemed very promising with regard to the above assumption of a linear relationship between the number of data points and the time required for computation. In order to obtain more evidence, two further tests were carried out using data sets of 3,200 (80x40) and 6,000 (100x60) points, respectively. All the results are shown in Table 8.4. A graphical representation of these results is shown in Fig.8.3.

Table 8.4 Variation of required cpu time with number of points

	Test 1	Test 2	Test 3	Test 4
No. of Pts	1,862	3,200	6,000	8,588
Matrix Size	49 x 38	80 x 40	100 x 60	113 x 76
CPU Time	3,506ms	6,458ms	12,137ms	16,645ms
CPU Time/Pt	1.88ms	2.02ms	2.02ms	1.94ms

During these tests, it was also found that the values for the cpu times may be slightly (up to 4%) different if the same data set is run at different times within the multi-user mainframe computer environment. The figures for Tests 1 and 4 in Table 8.4 were obtained at the same

time while the values for Tests 2 and 3 were obtained at another time. Therefore, from these test results, the assumption of the linear relationship between the number of data points and the cpu time required for processing these data points is acceptable. Furthermore, it might also be concluded that 2ms per point is the approximate time for this smoothing processing.

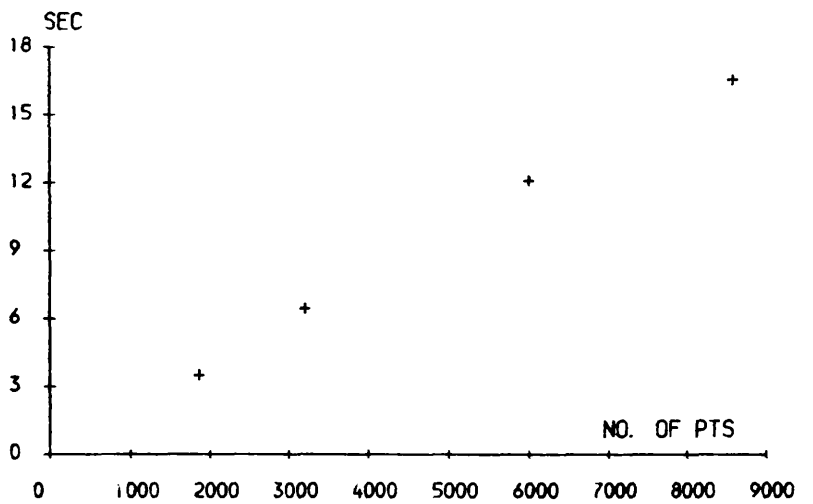


Fig.8.3 Relationship between required cpu time and no. of data points

(2). Estimation of the cpu time required for an entire model

Another point regarding the computational aspect which is of interest is the cpu time needed for smoothing the data set for an entire stereo-model. Since, the cpu time required for a data set is proportional to the number of data points, it is necessary, first of all, to estimate the number of points which will be measured in an entire stereo-model.

As is well-known in photogrammetry, a 60% forward overlap is normally used to form individual stereo-models. Also, in most photogrammetric practice, those areas about 1.5cm from each edge of the photo (the sidelap areas), are not used. Therefore the area of a stereo-model at photo scale is 20cm x 9.2cm. For such an area, the possible number of data points using different pixel sizes is listed in Table 8.5.

Supposing that every point is measured (- this is possible by automatic digital image correlation using the multipoint matching technique -), then the cpu times required to smooth the whole data set can also be estimated. Some examples are also shown in Table 8.5.

Table 8.5 CPU times required to smooth the whole data set from an entire model with different pixel sizes

Pixel Size	Possible No. of Data Points	Required CPU Time
32 um	2,875 x 6,250 = 17,968,750	35,937,000ms = 9.98hr
25 um	3,680 x 8,000 = 29,440,000	58,880,000ms = 16.36hr
12 um	7,666 x 16,666 = 127,761,000	255,520,000ms = 70.98hr
10 um	9,200 x 20,000 = 184,000,000	368,000,000ms =102.22hr

The results listed above are extrapolated from the limited test results. Therefore, they might not be very reliable. Nevertheless they do provide some very general information about the amount of computation required to smooth the data set for an entire model, with every point measured. The range of the computational times ranging from 35,973secs (=9.98hrs) to 368,000ms (=102.22hrs) is very substantial, given the power of the ICL 3980 mainframe machine, and certainly will cause many practitioners to think hard and long before deciding to implement such filtering techniques.

8.4 An algorithm for detecting gross errors in a gridded data set

8.4.1 Introduction

Often the presence of gross errors will distort the image (i.e. the appearance) of the spatial variation present in digital elevation data sets much more seriously than that resulting from random noise. In some cases, totally undesirable and even unacceptable results may be produced in the DTM and in the products derived from it due to the existence of such gross errors. Therefore, some methods need to be devised to detect this type of error in a digital elevation data set and then to ensure their removal from the data set.

As discussed before, DTM source data may be either in the form of a regular grid or it could be irregularly distributed. Regularly gridded data has certain special characteristics, e.g. it can be stored in the concise and economic form of a height matrix. These special characteristics can also be of help in designing an algorithm for gross error detection in such a data set. However, an algorithm which is suitable for application to gridded data is very unlikely to suit irregularly distributed data. Therefore, a quite different approach needs to be taken for the detection of gross errors in each of these two cases.

In this section, an algorithm for the detection of gross errors in a regularly gridded data set will be described, and some experimental results using this algorithm will also be presented. The algorithms for detecting gross errors in irregularly distributed data will be presented later in Section 8.5

8.4.2 Theoretical background to the algorithm development

The first question arising in algorithm development is "what kind of information can be used for this purpose"? To answer this question, it is necessary to refer back to the discussions in Chapter 3, where, the view is expressed that **slope** is the fundamental attribute of a surface point. Therefore, the answer to the question posed above is that, in this case, slope information can probably serve as the basis for the development of a suitable algorithm.

If this first point is accepted, then the second problem to be considered is the feasibility of obtaining the **slope information** from the set of gridded data. It is clear that the computation of the slopes existing at each individual grid point in different directions is reasonably convenient to arrange, and does not present a real problem. In view of this, it appeared promising to make use of slope information as the basis of the algorithm for detecting gross errors in gridded data.

From inspection of the DTM literature, it was found that Hannah (1981) had already developed an algorithm for such a purpose. Therefore, it seems pertinent to have an insight into her algorithm to judge the merits and demerits inherent in this approach so that a more comprehensive algorithm might be developed.

The principle of **Hannah's algorithm** is briefly as follows: As a first step, the slopes between the point under investigation, say P, and its neighbours (eight if not located on the boundary) are computed. Once this has been done for the whole data set, then three sets of tests are carried out on the slopes.

i). The first of these is called a **slope constraining test**, which checks the (eight) slopes immediately surrounding P to see if they are reasonable, i.e. whether they exceed the pre-defined threshold value (an absolute value) or not.

ii). The second is called the **local neighbour slope consistency test**, which checks the four pairs of slopes crossing P to see if the absolute value of the difference in slope in each pair exceeds the given threshold value.

iii). The third is the **distant neighbour slope consistency test**, which is very similar to the second. This test checks whether pairs of slopes approaching a point across each of the eight neighbours are consistent or not.

The results of these three tests will be used as the basis to judge whether a point is accepted or rejected. The overall result of using this algorithm is that it produces oversmoothing in areas of rough terrain in order to detect the gross errors or other unnatural features in relatively smooth areas of terrain.

However, in many ways, the most serious demerit of Hannah's algorithm is that all the given criteria for acceptance or rejection of data are expressed in an absolute sense. Obviously, the absolute slope values and slope differences will vary from place to place. For example, in an area with rough terrain, the absolute slope differences would be larger than those found in relatively smooth areas. The absolute values of slopes in the steep areas will be much larger than those found in relatively flat areas. That is to say, it is impossible for an absolute threshold value to be suitable for an overall area of interest except in a very homogeneous area. This discussion suggests that some criteria should be defined that can be used in a relative sense rather than simply setting absolute values. This is the basic starting point for the development of the new algorithm.

8.4.3 The principle of detecting gross errors

(1). General principle

The algorithm which is to be described in this section is based on the concept of **slope consistency**. Instead of the absolute values of slope change, relative values are considered. Furthermore, a statistic is computed from these relative values which can be used as the threshold value to measure the validity of a data point instead of using a pre-defined value.

Essentially, there two main differences between the algorithm to be presented in this section and Hannah's algorithm. In this new algorithm, first of all, the **relative slope changes** instead of the absolute values are being considered; secondly, the criterion for acceptance or rejection of a specific height value will be based on the statistical information about these relative slope changes, instead of using a pre-defined absolute value. The principle of this algorithm is as follows:

As shown in Fig.8.4, a data point, say P, can be defined by its row and column values, (I,J) within the height matrix. Its eight immediate neighbours - points 5, 6, 7, 10, 12, 15, 16, and 17 - can also be

defined by the rows and columns as (I+1,J-1), (I+1,J), (I+1,J+1), (I, J-1), (I,J+1), (I-1,J-1), (I-1,J), and (I-1,J+1).

From these eight points and point P itself, six slopes can be computed in both the row (I) and column (J) directions. Taking the row direction as an example, six slopes - those between points 5 and 6, 6 and 7, 10 and P, P and 12, 15 and 16 as well as 16 and 17 can be computed. From each set of six slope values, three slope changes can then be computed. For example, the slope changes at points 6, P and 16 can be computed from those slope values mentioned above. As stated before, these initial values are given in an absolute sense and will vary from place to place. Therefore, some relative values need to be computed from them to serve the purpose of the test.

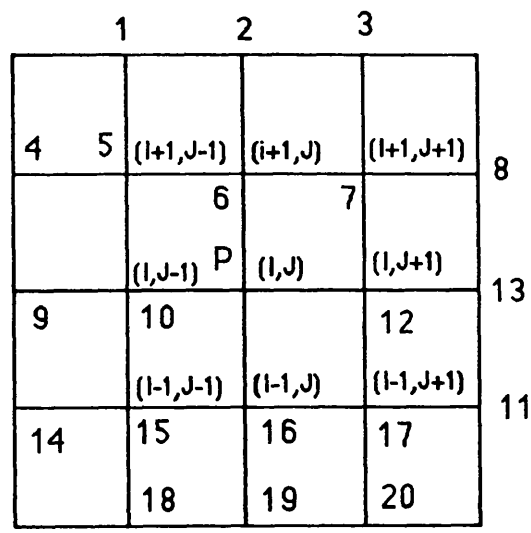


Fig.8.4 Point P and its neighbours

As one can imagine, if there is no gross error at point P, then for the same direction (e.g. the row direction), the **difference in the slope change (DSC)** at point P and that at its immediate neighbour (e.g. point 6 or point 16 located in the row direction) will be consistent, even though the absolute values of slope and slope change may vary from place to place. Therefore, these differences in slope change (DSC) are the relative values which are being searched for and can be used as the basis of the assessment of slope consistency and thus as the basis of a method of detecting gross errors.

That is to say, for every point except those along the boundary, two DSC values can be computed from the three slope changes in each direction. The DSC values from all the data points will be used as the basic information for this algorithm. From all these DSC values, a statistic will be computed and it will then be used to construct the required threshold value. Then this threshold value will be used as the

basis on which a judgement may be made as to whether a point has a gross error in elevation or not. For example, if all these four DSC values centred at a point, say P, exceed the threshold value, then point P will be regarded as having a gross error.

(2). Computation of the differences in slope changes (DSC values)

(a). The computation of the **slope** for the J direction, for example, is as follows:

$$SLOPEj(I+1,J-1) = (Z(I+1,J)-Z(I+1,J-1))/DIST(J-1,J) \qquad (8.4)$$

Where, DIST(J-1,J) is the distance between the points located at nodes (I+1,J) and (I+1,J). Similarly, the values of SLOPEj(I+1,J), SLOPEj(I,J-1), SLOPEj(I,J), SLOPEj(I-1,J-1) and SLOPEj(I-1,J) can be computed. The computation of slopes in the I direction will be implemented in a similar manner.

(b). After the computation of slopes, three **slope changes** in both directions can be computed as follows, e.g. in the J direction:

$$SLOPCj(I,J) = SLOPEj(I,J) - SLOPEj(I,J-1) \qquad (8.5)$$

Also SLOPCj(I+1,J) and SLOPCj(I-1,J) can be computed in a similar way. The computation of slope change in the I direction is also very similar.

(c). After this, two **differences in the slope changes** (the DSC values) for the point at (I,J) in each direction can be computed as follows:

J direction: $DSLOPCj(I,J,1) = SLOPCj(I,J) - SLOPCj(I+1,J)$

and: $DSLOPCj(I,J,2) = SLOPCj(I,J) - SLOPCj(I-1,J)$ (8.6)

I direction: $DSLOPCi(I,J,1) = SLOPCi(I,J) - SLOPCi(I,J-1)$

and: $DSLOPCi(I,J,2) = SLOPCi(I,J) - SLOPCi(I,J+1)$ (8.7)

All these values for the differences in slope change (DSC values) computed from all of the data points will be used for the computation of the threshold value for acceptance or rejection of data points.

Actually, the concept of computing slopes and slope differences is very similar to that used by Makarovic (1973) in the progressive sampling mrthod to compute the first height differences and second height differences. In his case, since square grids were used, the planimetric

structure of the data is homogeneous, therefore, the first and second differences in height can provide all the information required.

(3). Computing a threshold value

The consistency in slope changes can be validated through the use of some kind of statistic computed from the DSC values calculated from all the data points. For example, the absolute mean, the range (biggest minus smallest), the mode, root mean square error, as well as the standard deviation and mathematical mean are all possible options.

In this test, the mathematical mean and the standard deviation were first considered since these statistical values have many advantages over the others as discussed by the author (Li, 1988a). In fact, for the particular test which will be described in Section 8.4.5, it is also the case that the mathematical mean of the DSC values is very small and, therefore, the root mean square error (RMSE) is equally valid and used in this study. In this case, the **threshold value** is K times RMSE, where K is a constant.

There are three possible ways to compute the required threshold value(s) for this algorithm:

i). One of them is to compute only one RMSE value from all the DSC values at all the data points and in all directions;

ii). Another possible way is to compute four RMSE values from the DSC values at all the data points, one for each of the four sides (above, left, below, and right) defining each data point;

iii). The third possible way is to compute two RMSEs, one of which is related to the row (i.e. the I) direction and the other of which is for the column (i.e. the J) direction. In this case, the two DSC values of every point in the same direction, say the J direction, can also be added together to become a new value and the RMSE value can be computed from these new values.

Theoretically, the last method is the most reasonable because the absolute value of a sum of the two DSC values at the same point (e.g. P) in the same direction (e.g. the J direction) will become smaller (approaching zero) if the slope change is consistent, and it will become larger if it is not consistent. In this study, different criteria were tried and the final result proved this point. Therefore, a procedure based on the last method was implemented in this particular test study.

(4) Suspecting a point

All the methodology described above is designed to allow a judgement to be made whether a point has a gross error or not. A particular threshold value for an individual direction (e.g. the J direction) is used as the basis for a judgement to be made as whether the data points in this particular direction are acceptable or not. If the threshold value is exceeded at a point, then this point is regarded as being unnatural in the neighbourhood. In this case, it can be said that this point is **suspected** of having a gross error in this direction.

Actually, the procedures used for suspecting a point of having a gross error in the all these methods described above should be very similar since the only difference is to compare the DSC values with the overall RMSE or a particular RMSE value. Taking the second method as an example, if the absolute value of the difference in the slope change at a point along a single side is greater than the threshold value - K times the RMSE of this side - then the point is suspected of being unnatural in the context of the values in the neighbourhood of this point. If all four sides around the point are suspect, then, this particular point will certainly be suspected of having a gross error. In most cases, if three sides of a point exhibit characteristics which cause them to be suspect, then again it may then be regarded as having a gross error. For the last method, if a point is suspect in both the row and column direcions, then it is certainly regarded as having a gross error.

Another question which needs to be answered is what value of K should be adopted in each case. For different cases, different values may be used. In this specific study, it was found that the DSC values for the test area are quite normally distributed, thus a value of 3.0 was used.

8.4.4 Data correction

It is also possible that some gross errors may have not been detected in a single run if the gross errors are located closely together, in which case, a point which has a gross error may still be considered as natural if its neighbours also have gross errors of a similar magnitude. This means that a further detection of the remaining gross errors may be necessary in some cases. However, in this algorithm, since all data points are used for the computation of slopes and thus for the slope changes, correction of those points found to have gross errors must be carried out to secure an improvement in the results. Therefore, in order to have reliable values, on the basis of which the next run to detect gross error will take place, those data points which were regarded as having gross errors in the previous run need to be "corrected" immediately. The principle of data correction used in this

algorithm is as follows:

In Fig.8.4, suppose that point P is the point which is suspected of having a gross error, and points 1 to 20 are its neighbours. In the process of detecting the gross errors, the slope and slope change values at all these points have been calculated (except those points near boundaries). From points 6, 16, 10 and 12, four estimates have been made. The estimation from Point 10 may be taken as an example. The average value of the slope change values at points 5 and 15 (in the J direction) are taken as the estimated slope change at point 10 (in the same direction). The new slope at point 10 (to P) can then be computed as follows:

$$SLOPE(10,J) = SLOPE(9,J) + (SLOPC(5,J)+SLOPC(15,J))/2.0 \qquad (8.8)$$

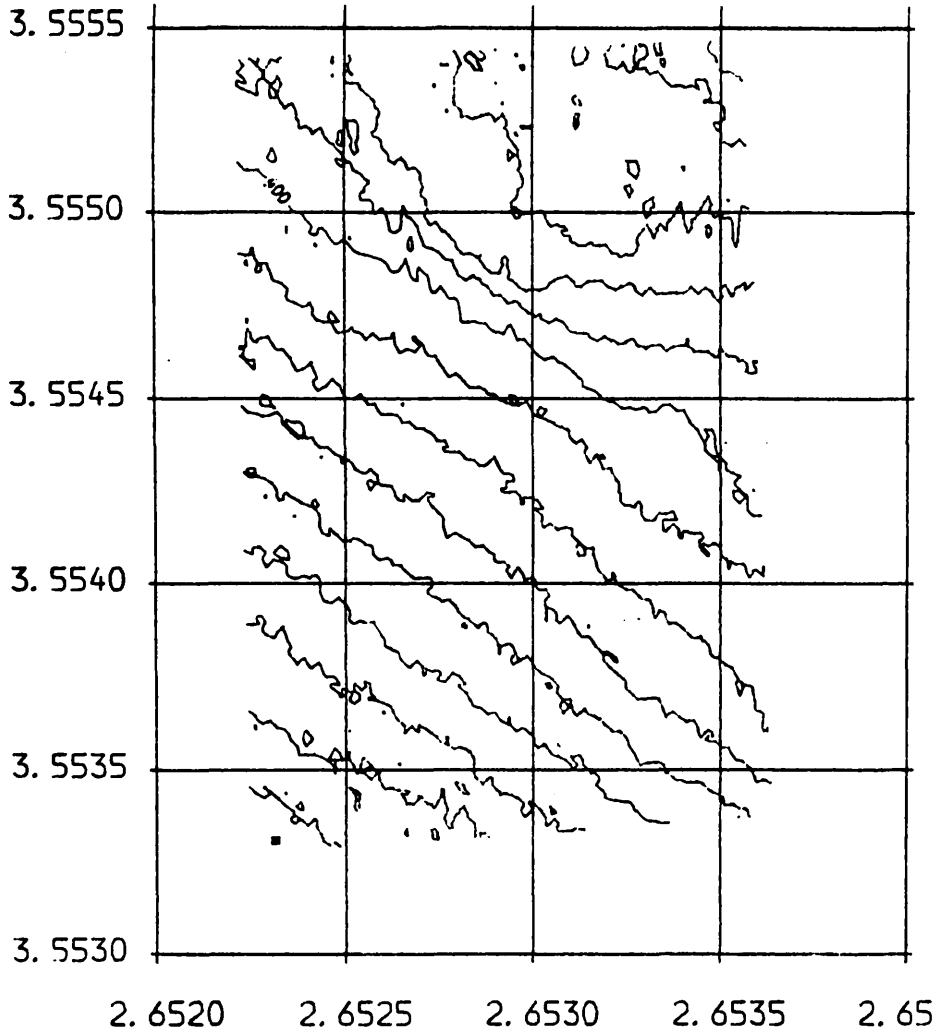
Where, SLOPE(10,J) denotes the slope at point 10 in the J direction; and SLOPC(15,J) denotes the slope change at point 15 in the J direction; and the definitions of the other terms in the equation will be similar to those.

This slope value is used to compute the height value for point P. At last, the average of four such estimates (from the four directions, i.e. above, left, below and right) is used as the height estimate for point P. Of course, if either point 9 or point 10 is also suspected of having a gross error, or if the other neighbours in this side (points 4, 5, 6, 14, 15, and 16) are suspected of having gross errors, then any estimate from this side will be unreliable and should not be used. Similarly, it is also possible that no reliable estimate can be made for point P in a single run. Therefore, some form of interactive processing is needed.

This algorithm can only be used to correct those suspected points which are not located near the boundary of the area covered by the data set. For those points along the boundary, no correction will be made.

8.4.5 Experimental validation and discussion

As mentioned at the end of Section 8.3, some unnatural features still seemed to be present along some of the contours produced from the data set after filtering. For example, in Fig.8.2b, the contours plotted for the height values of 330m, 430m, and 440m all exhibit unnatural characteristics. There are also two very tiny closed contours located in between the contours at 410m and 420m. It was then assumed that there may be some small gross errors, or at least some unnatural data points present in the data set. Therefore, this data set was used to test the algorithm presented in this section since no better data sets were available to the author.



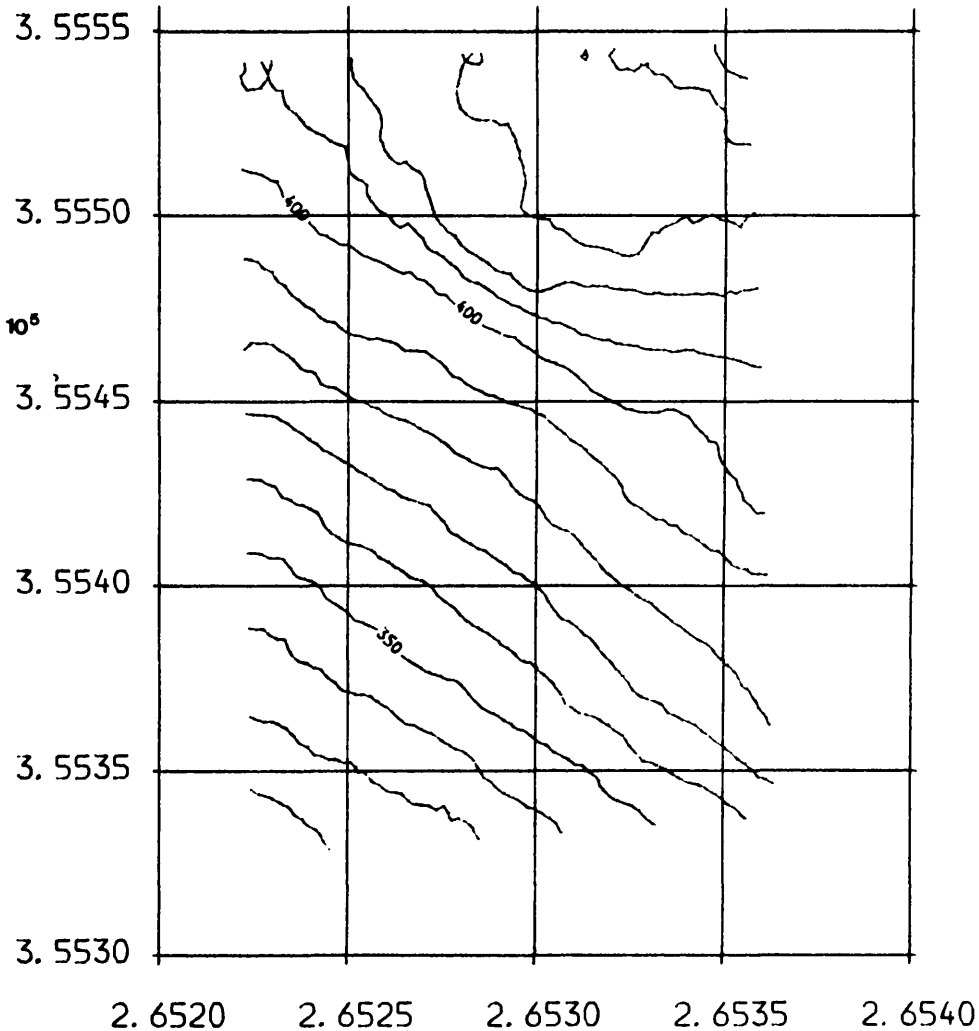


Fig.8.5 Quality improvement of the contours after gross error removal
(a). Contours produced from the original data set;
(b). Contours produced from the data set after smoothing and gross error removal

After applying the algorithm to the data set, these unnatural features have in fact been removed. The result is shown in Fig.8.5(b). For comparison, Fig.8.2(a) (the contour plot produced from the original data set) has been placed as Fig.8.5(a) alongside the plot after removal of gross errors.

As already mentioned above, for those points along and near the boundaries of the area, no error detection and correction can be applied. This is the reason why there are still some unnatural features occuring in the contours near the boundaries, especially along the top of the plot. Although the test carried out on this algorithm is very limited, yet the results show that it works quite well.

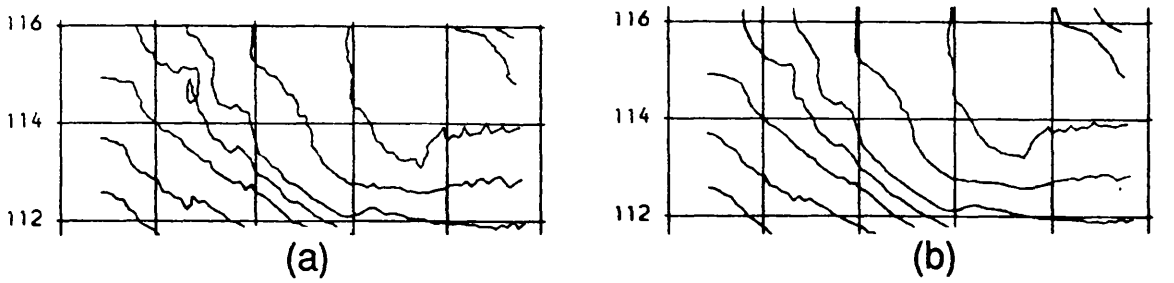


Fig.8.6 Another example of contours produced from the data set (a). before; and (b). after removal of gross error

Fig.8.6 is another example showing how this algorithm works. Fig.8.6(a) shows the contours produced from part of the data set comprising model coordinates of the same area after smoothing. Clearly, some residual errors exist producing unnatural features in some of these contours. After applying the gross error detection procedure, the corresponding contour plot (Fig.8.6b) shows that that these unnatural features have again been removed.

8.4.6 Discussion of the computational effort required to detect gross error in a regular gridded data set

As was done in the previous section (8.3.6), a brief discussion regarding the computational aspects of the gross error detection method is also desirable. The cpu time spent on the detection of gross error in this test data set of 8,588 points was 20,558ms which is a little more than the time - 16,645ms - spent on smoothing the same data set. It is equivalent to 2.39ms/point.

It might also be assumed that the amount of cpu time required to detect the gross errors in a regularly gridded data set is proportional to the number of data points since the amount of computation for every data point should be the same. If this assumption applies, then the amount of cpu time needed for processing the whole data set from an entire

model is listed in Table 8.6.

It must be said that, once again, this shows that an intensive computational effort needs to be made if gross error detection is to be applied to a dense gridded data set. Even with a powerful computation engine such as the ICL 3980, the range of times required for such a single stereo-model varies between 43,014secs (=11.95hrs) to 440,460secs (=122.35hrs). Again, if applied, these are very expensive overheads to be added to the DTM process and it is an open question whether potential users will feel that the benefits justify the very considerable expense needed to improve the quality of the DTM.

Table 8.6 CPU times required to detect gross errors in the gridded data set from an entire stereo-model

Pixel Size	Possible No. of Data Points	Required CPU Time
32 um	2875 x 6250 = 17,968,750	43,014,000ms = 11.95hr
25 um	3680 x 8000 = 29,440,000	70,474,000ms = 19.58hr
12 um	7666 x 16666 = 127,761,000	305,835,000ms = 84.95hr
10 um	9200 x 20000 = 184,000,000	440,460,000ms =122.35hr

Actually, it may not be necessary to apply this algorithm to the whole data set. Sometimes, a visual inspection of the perspective plots or contours produced from the data set may help in finding where the gross errors are located. In this case, this algorithm need only be applied to those areas where gross errors exist. In this case, the cpu time required for processing should be much smaller than that listed in Table 8.6.

8.5 Algorithms for detecting gross errors in an irregularly distributed data set

8.5.1 Introduction

The algorithm described in the previous section is based on the consistency of slope changes in the neighbourhood of a point under investigation within a regular grid-based data set. However, if the data points are irregularly distributed, then difficulties in checking the consistency of slope change will be met. It will, therefore, not be applicable in this case.

In an irregularly distributed data set, the information which is conveniently available to users is the set of X,Y,Z coordinates of the

data points. Therefore, in this case, the height information for every data point and its neighbours can still be used as the basis on which to assess the validity of the data point elevations. The algorithm to be described in this section is based on this height information.

Gross errors may be scattered as isolated occurrences in the data set or they might occur in clusters. In the latter case, the situation is much more complicated. So, in this chapter, an algorithm for detecting the individual gross errors scattered in the data set will first be described. Then this algorithm will be modified to suit the detection of gross errors occuring in clusters. Some experimental results will also be presented.

8.5.2 Approaches for algorithm development

Depending on the size of the area to be considered, three approaches can be distinguished to develop an algorithm for gross error detection, namely, a global approach; a regional approach and a local approach. In this sense, it has some similarity to the basis on which interpolation methods were discussed in Chapter 7.

Any method using a **global approach** must involve the construction of a global surface through all points in the data set using a high-order polynomial function, then checking the deviation of every data point from the constructed surface. If the deviation at a point is greater than the threshold value, then this point can be considered as having a gross error. The threshold value might be a pre-defined one or it may be computed from the deviations of the heights of the data points from the global surface. As Hannah (1981) pointed out, global techniques "have the drawback that they give identical treatment to all areas". However, terrain is rarely uniform in roughness, so the uniform application of a global technique to an area may result in too many points being regarded as having gross errors in rough areas, which in fact are not errors, while failing to detect gross errors in relatively smooth areas. That is to say, the final result might be totally undesirable or indeed wrong.

The methods employed in a **regional approach** could be very similar to those used in the global approach, i.e. constructing a regional surface and then checking the deviations of the data points from the specific surface. The only difference is the size of the area which the surface covers. The adequacy of this approach depends partly on the size of the area which a particular regional surface covers.

A major **drawback** to the method of using a polynomial surface, regardless of the size of the area which such a surface covers, is that those points which have gross errors will also have been used to

construct the DTM surface. In this case, all points near the particular point with a very large gross error may have large deviations from the constructed DTM surface due to the large influence of the erroneous point on the constructed surface. Thus they may all have been identified as having gross errors when in fact this is not the case.

If a **local approach** is employed, then the use of a polynomial surface to fit the data points can be avoided. A method similar to that used in pointwise interpolation, and consequently referred to as a **pointwise method** in this context, can be employed. This method involves comparing the height value of the point under investigation with a representative value such as the average height derived from the heights of its neighbours. As a result, if the difference is larger than a certain threshold value, then this point can be regarded as having a gross error (see Section 8.5.3).

The principle of the pointwise method is so simple and intuitive and the computation could also be so simple (e.g. simply compute the average) that a procedure based on this method has been developed and will be presented below.

8.5.3 An algorithm based on the pointwise method (Algorithm 1)

(1). General principle

The general principle is as follows. For a specific point which is being tested, say P, a window of a certain size is first defined centred on P. Then a representative value will be computed from all the points located within this window. This computed value is then regarded as an appropriate estimate for the height value of the point, in this case, P. Or this value can be regarded as the "true value" of point P. By comparing the measured value (of P) with the representative value estimated from the neighbours, a difference in height can be obtained. If this difference is larger than the computed threshold value, then this point can be considered as having a gross error. The computation of threshold values will be discussed later.

In this method, the height value of the point P which is being tested is not taken into consideration when computing the representative value for P. Therefore, the height value of point P has no influence on the estimated value derived from the neighbours. Thus the height difference obtained through this procedure provides more reliable information about the relationship between point P and its neighbours.

(2). Range of neighbours

The range of the area within which neighbouring points will be searched

for is (like that used in the pointwise interpolation method) specified by a window centred on point P. This can be specified by defining either an area, or the number of nearest neighbours required. The former can be expressed as follows:

$$\begin{aligned} &\text{X range:} && X_p - Dx < X_i < X_p + Dx \\ &\text{Y range:} && Y_p - Dy < Y_i < Y_p + Dy \end{aligned} \qquad (8.9)$$

Where, X_p and Y_p are the X and Y coordinates of P - the point under inspection; X_i and Y_i are the X and Y coordinates of the ith point in the neighbourhood; and Dx and Dy are the half-window sizes in the X and Y directions respectively.

Also, a combination of both criteria can be used. The average window size can be computed according to the total number of points and the coordinate ranges of the area. This average value can be used as the initial window size. In an area with a higher density of points, the number of points lying within a window of this size will be larger than average. However, in a lower density area, it may happen that only a few points are located inside such a window. Therefore, a minimum number of points may also need to be specified. If the number of points within a window is smaller than the specified value, then the window is enlarged a little until the specified number of points is reached.

(3). Computation of the representative value

In this algorithm, the average of the height values of the neighbours is used as the representative value. The average (height) value can be computed in either of two ways. The simpler one is to simply take the **arithmetic mean** and the other is to use a weight for every point according to its distance to the point under investigation, e.g. the weight is inversely proportional to the distance.

The **weighted mean** should be closer to the real value of the central point - P in this example - if there are no gross errors in the neighbourhood. However, if a point with a large gross error is very close to the central point (P), then the weighted mean is greatly affected by this point, thus producing a very unreliable value. Therefore, the simple arithmetic mean may be more desirable. In fact, practical tests confirmed this particular point. Since the calculation of the simple arithmetic mean will take much less computation time, it has, therefore, been used in this algorithm.

(4). Computing threshold value and suspecting a point

The height differences of all points are used to compute a statistical

value which will then serve as the basis for deciding on a threshold value.

Suppose, M_i is the arithmetic mean of the neighbouring heights centred at the ith point in the data set and the difference between the M_i and the height value of this (ith) point (H_i) is V_i, then

$$V_i = H_i - M_i \qquad (8.10)$$

If the data set has N points, then the total number of V values is also N. The required statistical value can then be computed from these values of v. In this study, the mathematical mean (u) and standard deviation (SD) are computed from these values of V and are then used as the basis for constructing the threshold value. A threshold value, KxSD, which is similar to the one discussed in the previous section, is used in this algorithm with K=3. After the threshold value has been set, every data point in this data set can be checked. For any point i, if the absolute value of (V_i-u) is larger than this threshold value, then this point, i in this example, is considered as having a gross error.

8.5.4 Experimental tests using Algorithm 1

Two sets of irregularly distributed data were available to the author and were used for this study. The data sets have been compiled from digital image correlation in a DSP and were again kindly provided by my research colleague, Mr. Ali Azizi (1990).

(1). Results obtained from the first data set

The distribution of the first data set and the contours produced from this set of data are shown in Fig.8.7. Fig.8.7(a) shows the irregular distribution of the data points. Fig.8.7(b) (the corresponding contour plot) shows clearly that there are some gross errors in the data set, which need to be detected. The size of this area is about 4.5cm x 4.5cm on the photo and about 800m x 800m on the ground respectively. Within this area, the height values of 4,964 points were correlated.

The simple arithmetic mean was used as the representative value derived from the neighbouring points while the window size was defined by the combination of specifying an area size and a certain number of points. The minimum number of points was initially specified as five. As a result, the algorithm did not work well. Then the number was gradually increased. For this data set, a number lying between 15 and 20 gave the best results.

After applying this algorithm, those points which created the unnatural

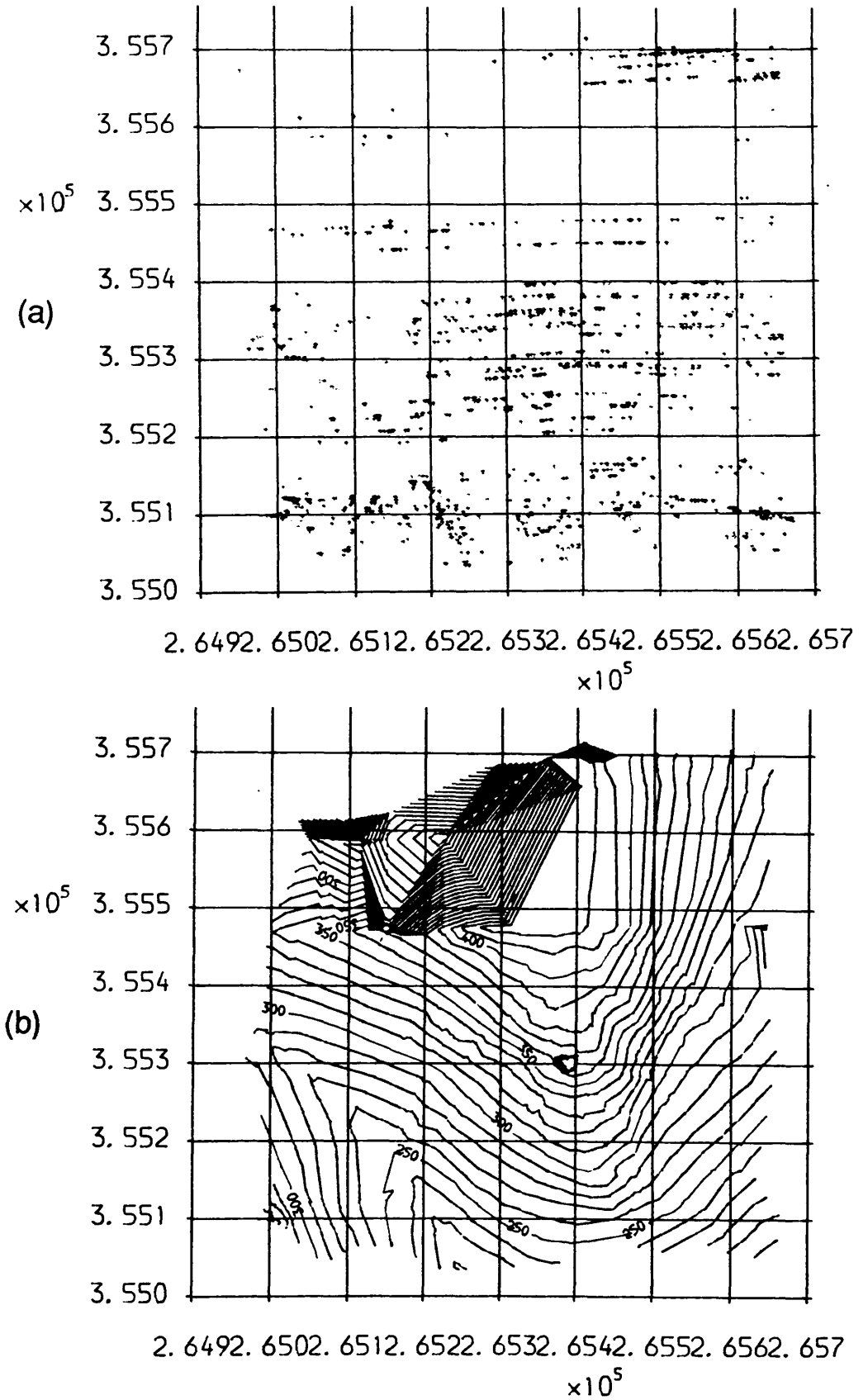
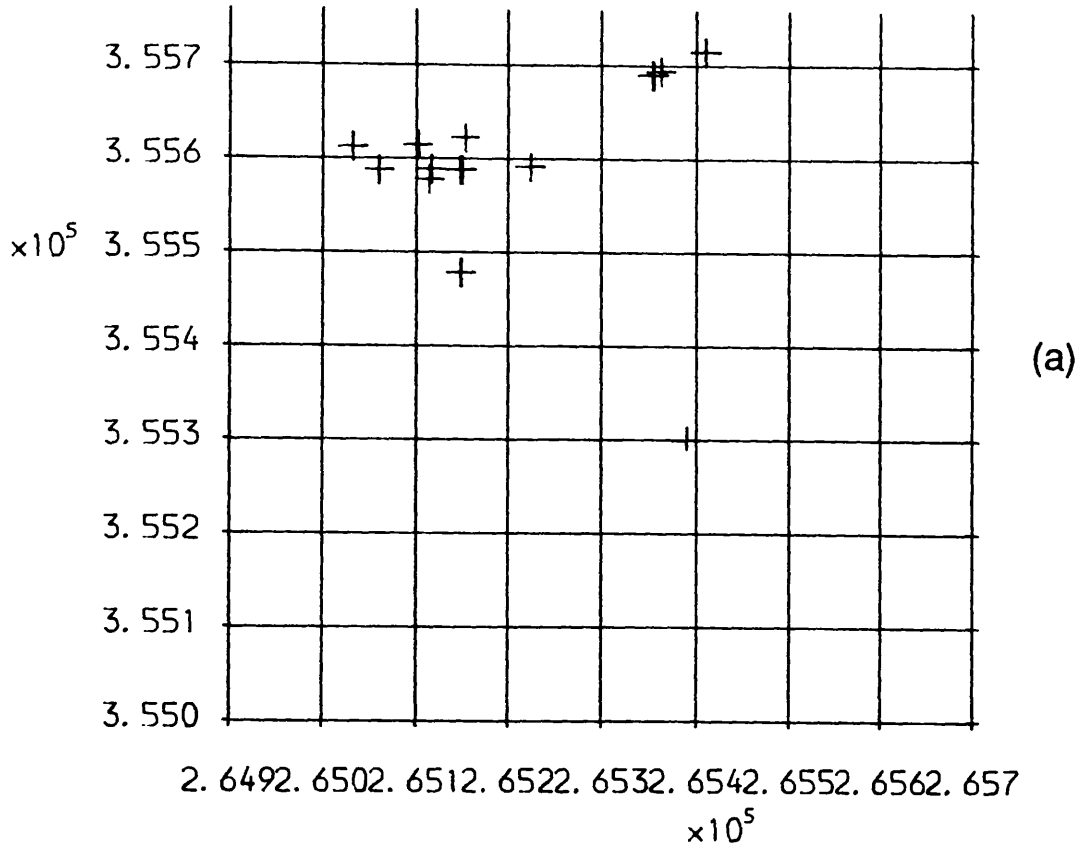


Fig.8.7 Information about the first data set
(a). Data distribution:
(b). Contour plot of the data set
which indicates gross errors



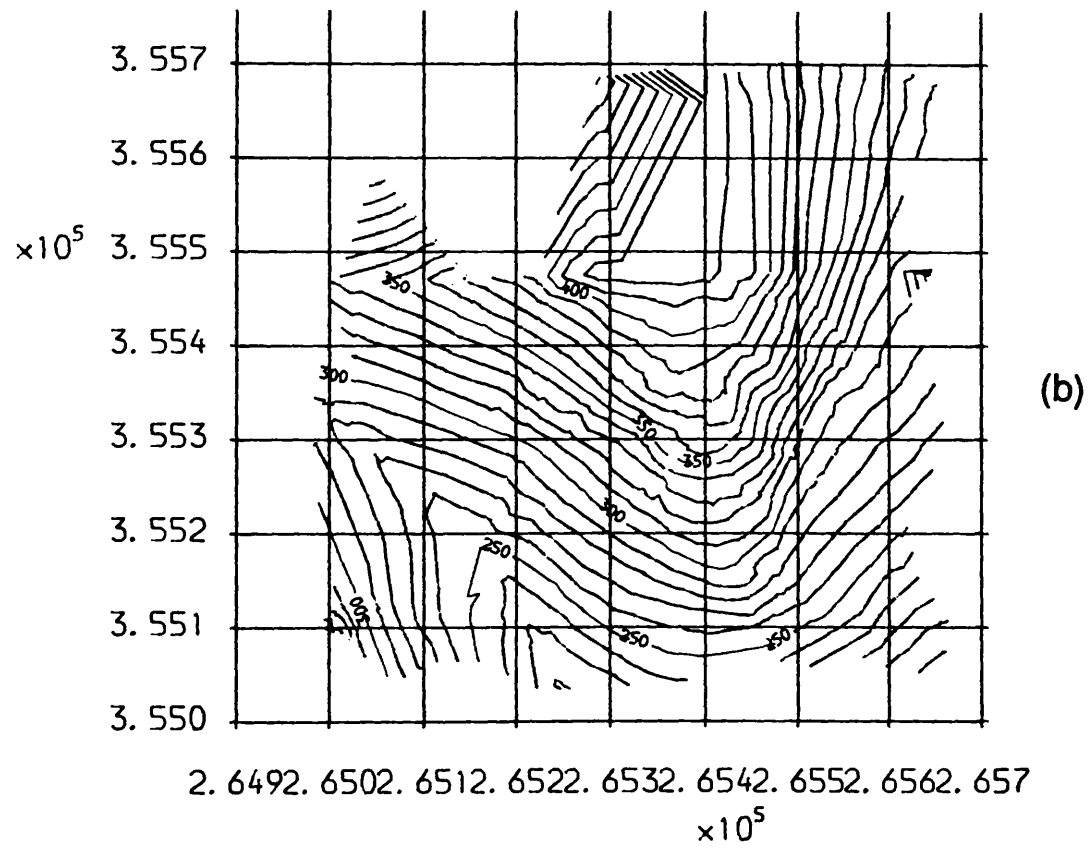
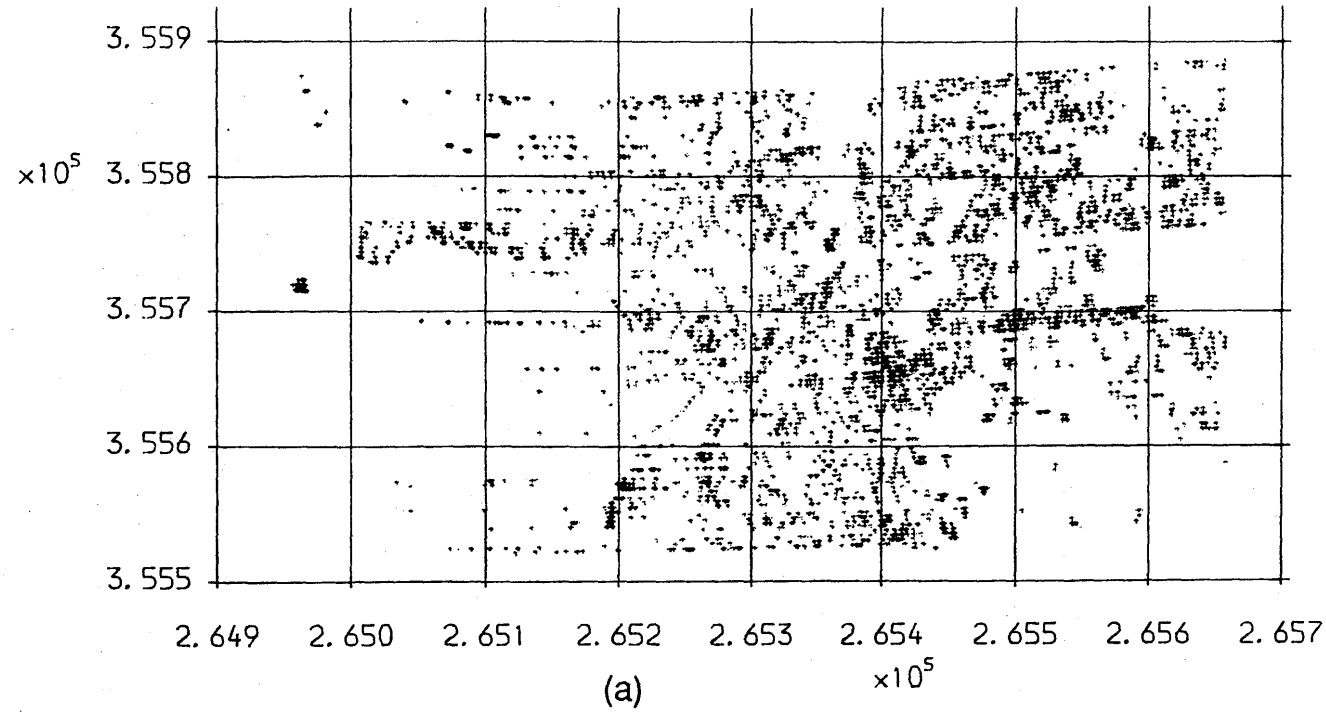


Fig.8.8 Results obtained by Algorithm 1 for Area 1
(a). Gross errors detected by Algorithm 1;
(b). Contour plot from the data set after removal of gross errors



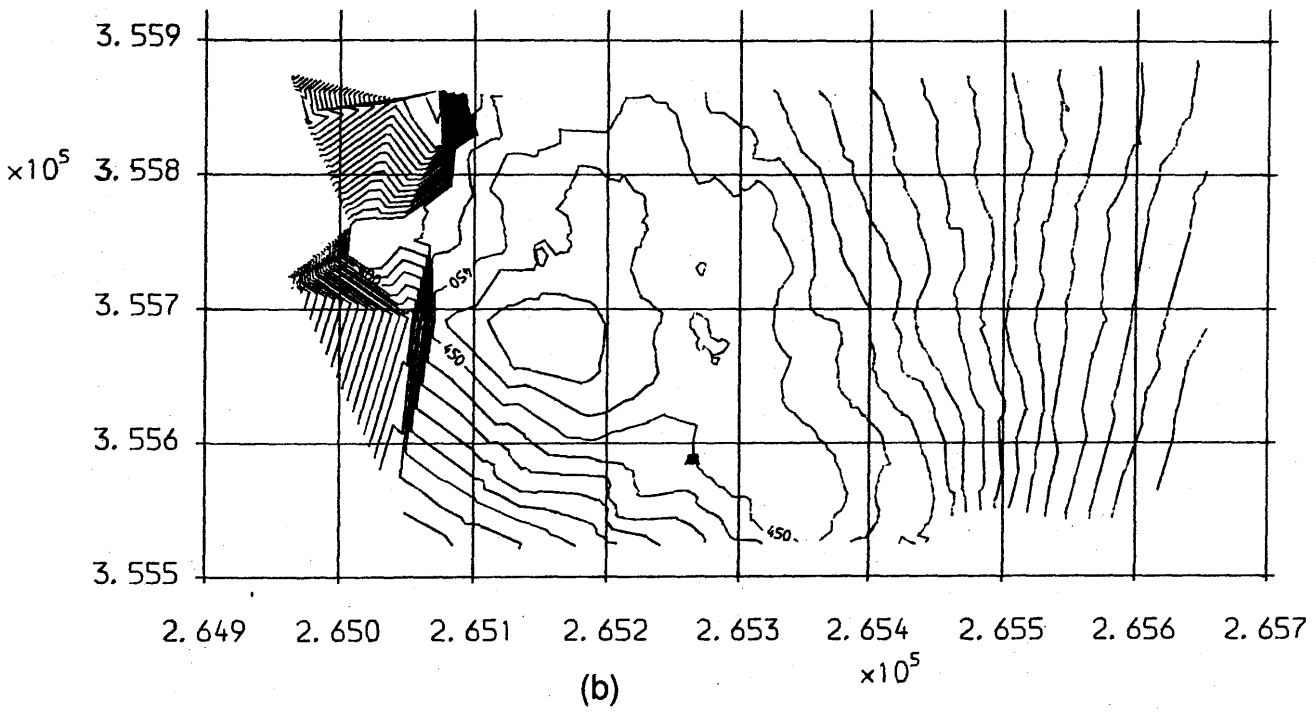
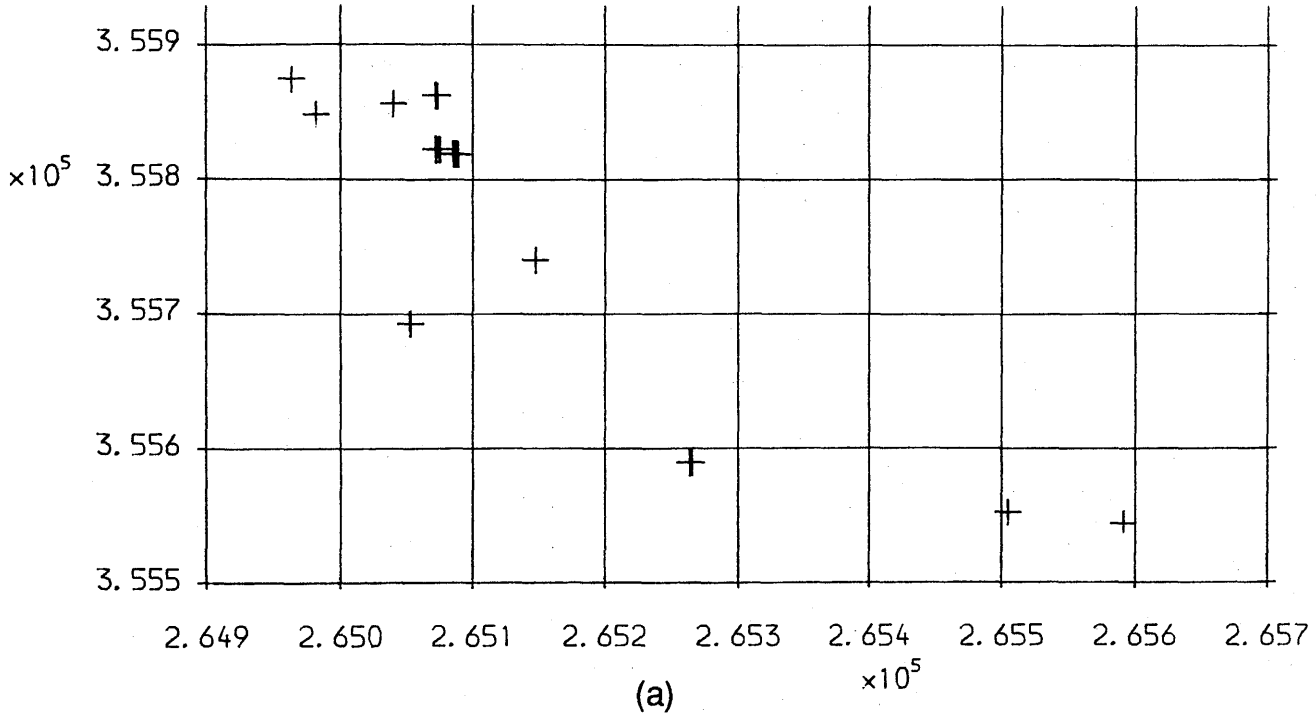


Fig.8.9 Information about the second data set
(a). Data distribution;
(b). Contour plot of the data set which indicates gross errors



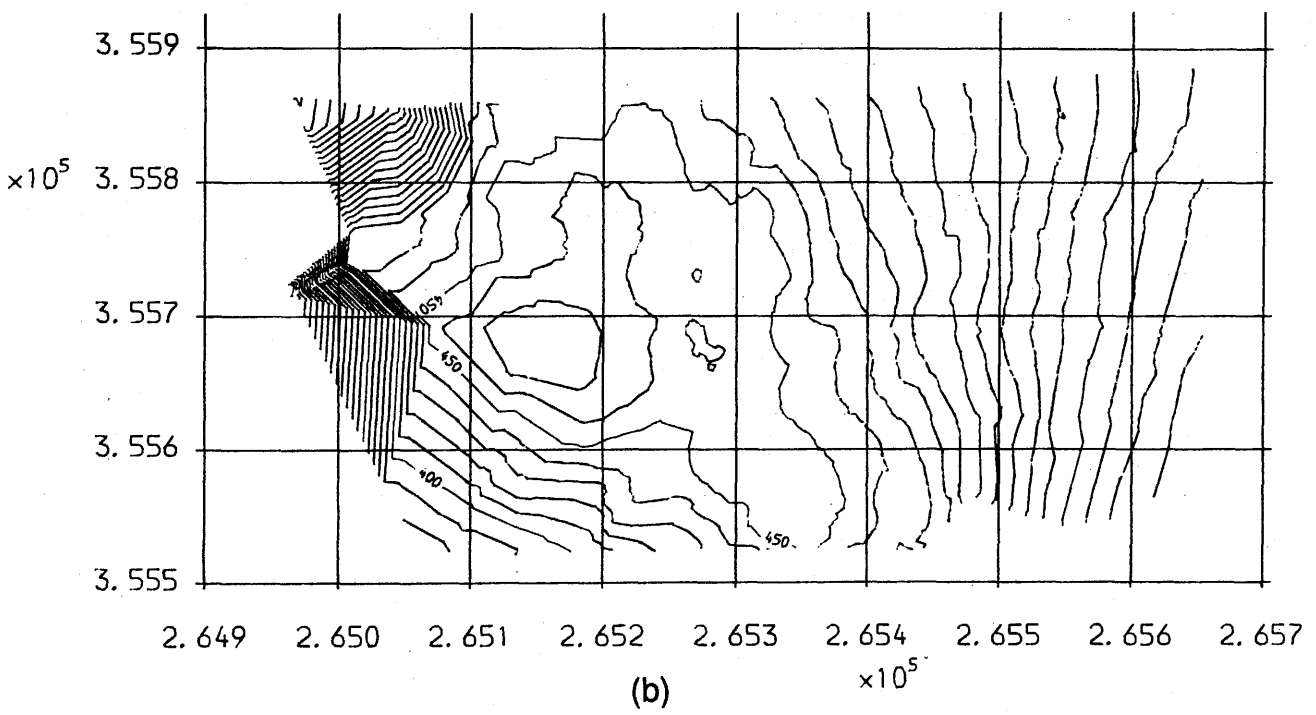


Fig.8.10 Results obtained by Algorithm 1 for Area 2
(a). Gross errors detected by Algorithm 1;
(b). Contour plot which indicates the further presence of gross errors

contours (see Fig.8.7b) were detected and their locations plotted in Fig.8.8(a). The distribution of the data points after removal of those erroneous points and the contours produced from this corresponding data set are shown in Fig.8.8(b). The results show that this algorithm works very well.

The search for the neighbours of every data point was undertaken over the whole of the test area. As a result, it took a very considerable amount of cpu time - 324,118ms - to complete this exercise for this data set with a number of 4,694 points. It is about 69ms/point.

(2). Results from the second data set

The distribution of the data points in the second data set and the contours produced from this data set are shown in Figs.8.9(a) and (b). The size of this test area is about 4.0cm x 2.2cm on the photo and about 700m x 400m on the ground. In total, the height values of 4,733 points were available within this area. Again, Fig.8.9(b) shows clearly that some gross errors are present in the data set. Once again, the author's algorithm was applied to this data set.

Initially, the same parameters and window size as those used for the previous test were used for this test. The detected points and the contours produced from the data after removal of the erroneous points are shown in Fig.8.10. From the contour plot (Fig.8.10b), it can be seen that there are still some gross errors left in the data set.

A much larger window size (containing 60 points) was also used, but still failed because the remaining gross errors occur in clusters and have a large magnitude as well. In order to solve this problem, another algorithm was developed which will be described below.

The search for the neighbours of every data point must be undertaken for the whole of the test area. It took a very considerable amount of cpu time - 403,906ms - to complete this exercise conducted on only 4,733 points. This amounts to about 85ms/point.

8.5.5 Algorithm for detecting a cluster of gross errors (Algorithm 2)

The algorithm described above appears to be suitable for the detection of scattered gross errors. To put the situation in another way, implicitly the assumption is made that there are serious gross errors in the window for every point. However, gross errors may be in a very tight cluster and with a large magnitude - indeed this could very well be the case where the data is derived using automatic correlation (image matching) techniques. In this case, this algorithm will fail to

detect these errors. Therefore, some further development or modifications need to be made to the algorithm presented above.

(1). A solution to the problem associated with Algorithm 1

Theoretically, the use of an increased window size should be a solution. However, as already described, it still failed to work with a window size of up to 60 points. If the window size is increased more and more, it might work in some cases, but the result may be unsatisfactory since the representative value derived from the neighbouring points may then be quite deviated from what it should be, thus leading to an unreliable judgement being made. Therefore, a search was made for an alternative solution.

The idea behind the algorithm development is to find those points which have a very great influence on the representative value, in this case, the average value. Then these points would not be taken into consideration while computing this representative value.

(2). Point data snooping in a window

The method used for point data snooping in a window is very similar to the idea used in Algorithm 1 for detecting gross errors. The procedure used is as follows:

First of all, the first point in the window area is taken out of the window and then a new value of the representative value - in this case, the average value - is computed from all the remaining points in the window; then the difference between this average value and the original one is computed and recorded. This procedure is then applied to every point in the window. Suppose, there are M points in the window, then M differences can be obtained as follows:

$$v_i = a_i - a \qquad (8.11)$$

Where, v_i is the difference between the new average, a_i, computed from all the remaining neighbouring points in the window other than the ith point and the original average value, a, which was computed from all the points in the window. The rest of the procudeure is the same was as done for detecting gross error in Algorithm 1. That is to say, the M values of v are used to compute a single statistical value, which is then used as the basis on which to construct a threshold value for snooping data points within the window. After that, every value of v can be checked. If any value of v, say v_j, exceeds this threshold value, then point J will be excluded from this window. In this way, all those points which appear to make a very great change in the representative value in a window can be excluded from this window.

This point data snooping technique is then applied to every window. After this has been done, the rest of the procedure is exactly the same as the procedures described before for Algorithm 1, i.e. computing a representative value, constructing a threshold value and identifying suspect points.

8.5.6 Experimental test using Algorithm 2

The second algorithm was then tested with the second data set. The errors detected by this algorithm are plotted in Fig.8.11(a) and the contours produced from the data set after removal of gross errors by Algorithm 2 are shown in Fig.8.11(b).

It can be seen clearly from Fig.8.11(b) that there is still a point with a small gross error located in the north-west of the test area since it produces an unnatural contour in that part of the plot. The reason why this point was not detected by this new version of the algorithm could be due to the fact that, in applying this algorithm, a larger window size needs to be used. In the case of this example, the minimum number of 35 points was specified. However, the use of a large window size results in a decrease in the sensitivity of this algorithm to gross error.

As expected, it took much more cpu time - 412,131ms - to run this program on the test data set. This amounts to about 87ms/point.

8.5.7 Discussion of algorithms used for the detection of gross error

From inspection of Figs.8.10 and 8.11, it can be found that the majority of the gross errors detected by these two programs are identical. However, each may omit one or more points for the specific reasons which have been discussed in previous sections. Therefore, a complementary use of both algorithms may produce desirable results, in which case, all the points detected by both of them should be deleted from the data set.

Fig.8.12(a) shows the gross errors detected by both programs while the contours produced from the data set after removal of these points are shown in Fig.8.12(b). It can be found that a much more reasonable result was produced after removal of the gross errors. (N.B., The unnatural contours at the bottom left corner of the plot are artificial since no data points are located there).

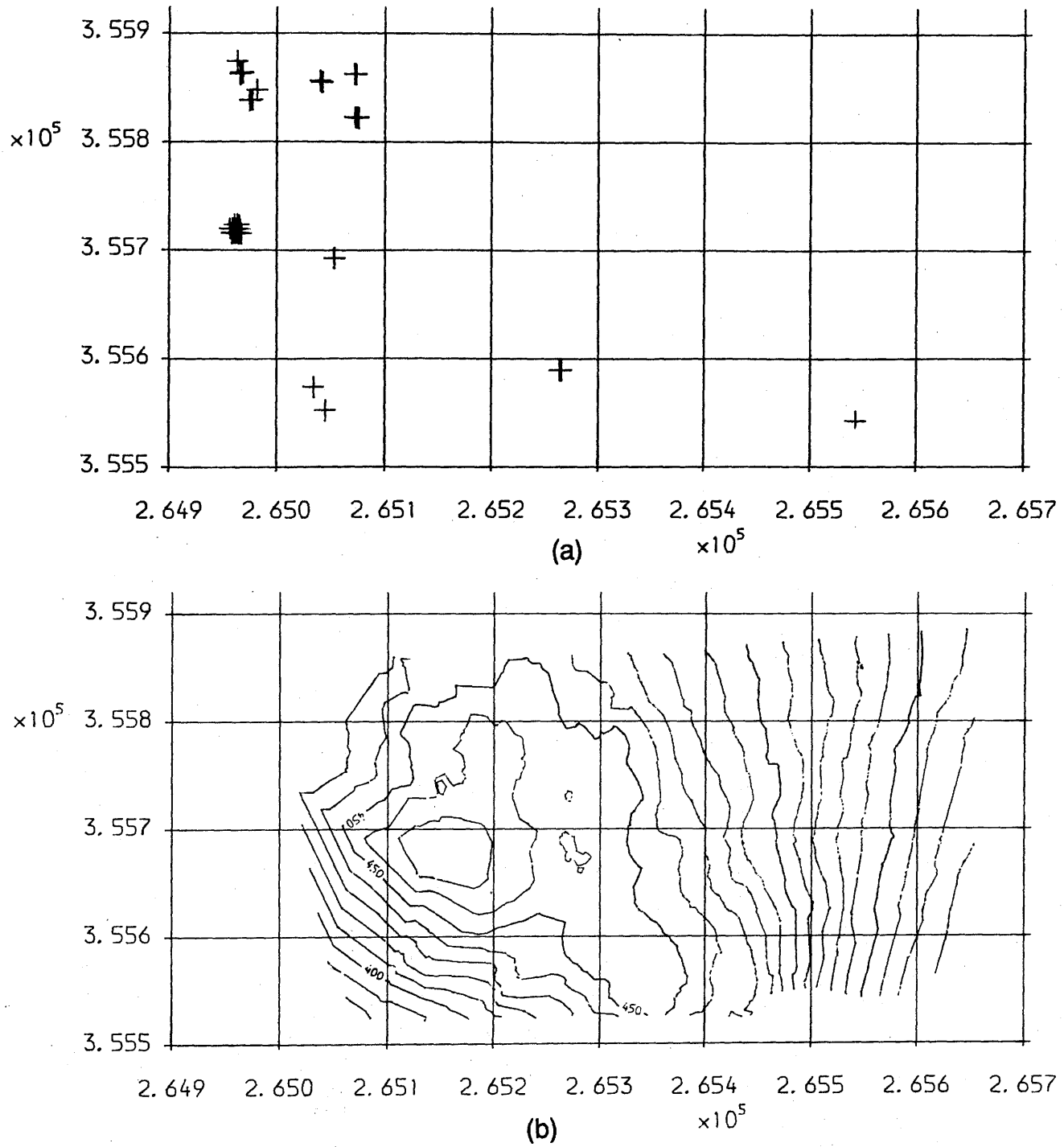
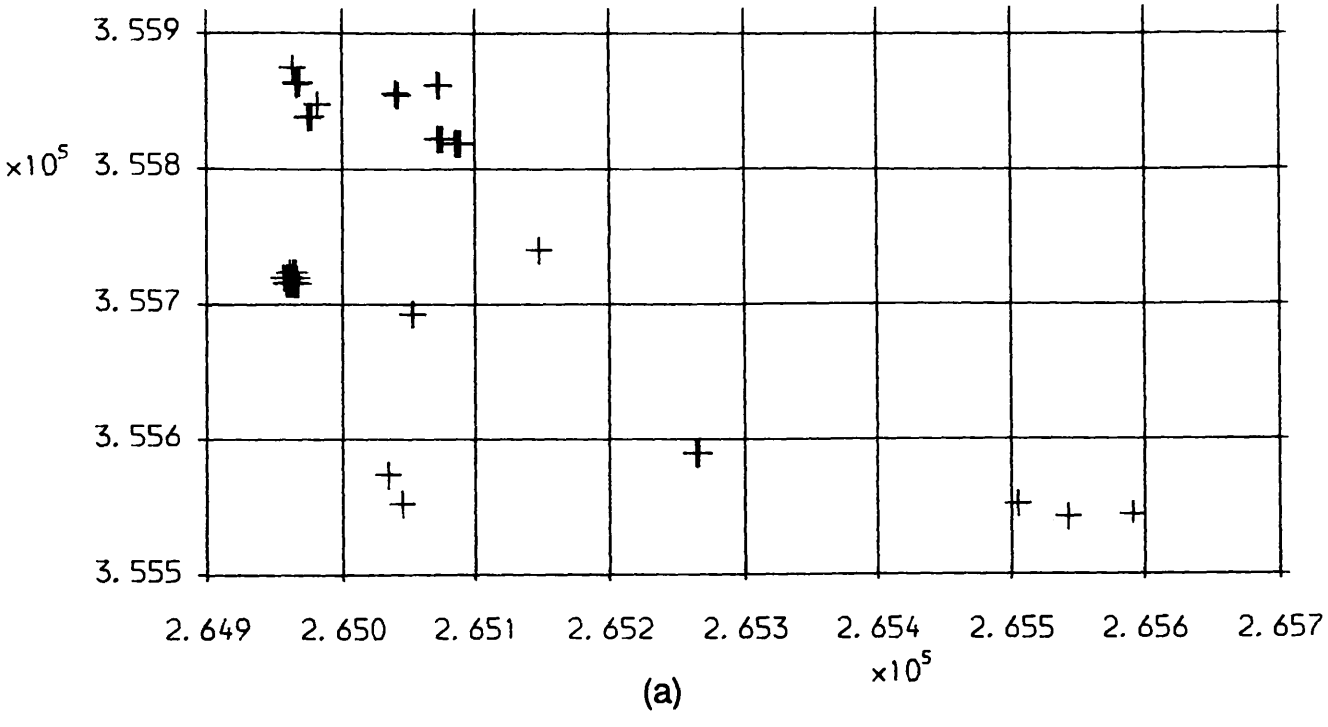


Fig.8.11 Results obtained by Algorithm 2 for Area 2
(a). Gross errors detected by Algorithm 2;
(b). Contour plot of the data set after gross error removal



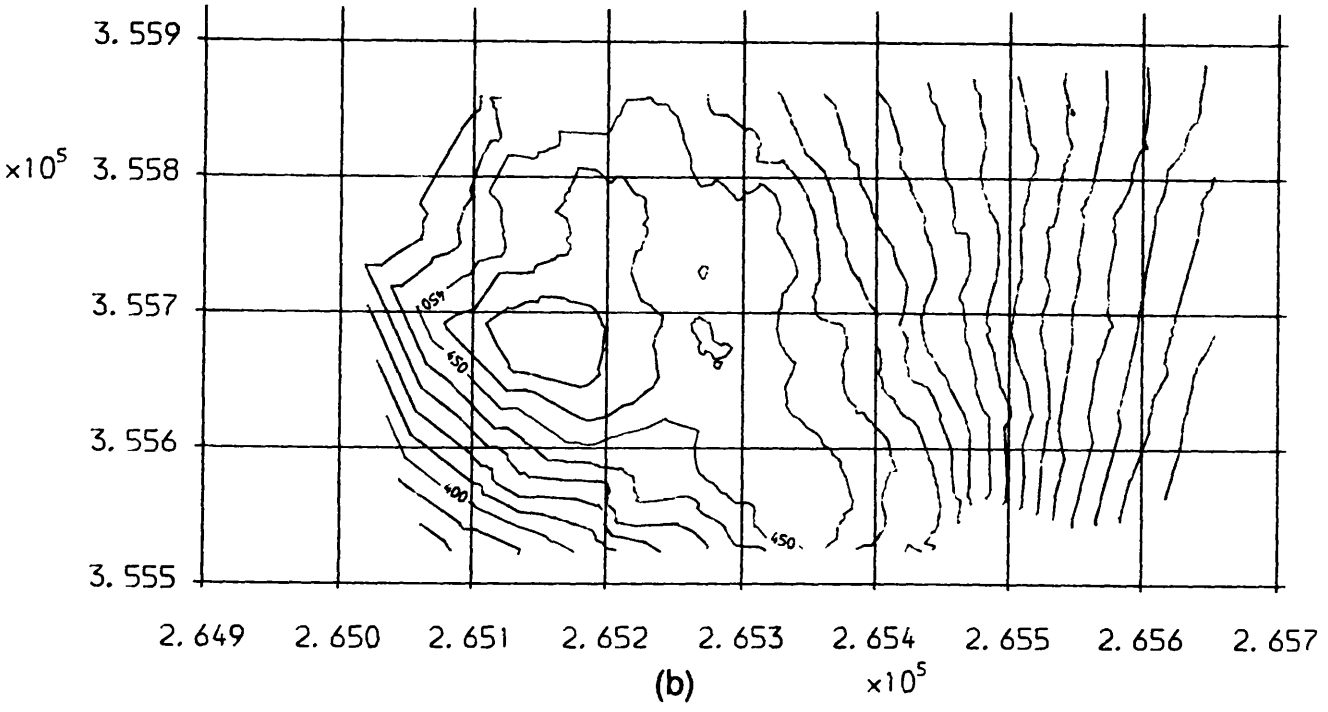
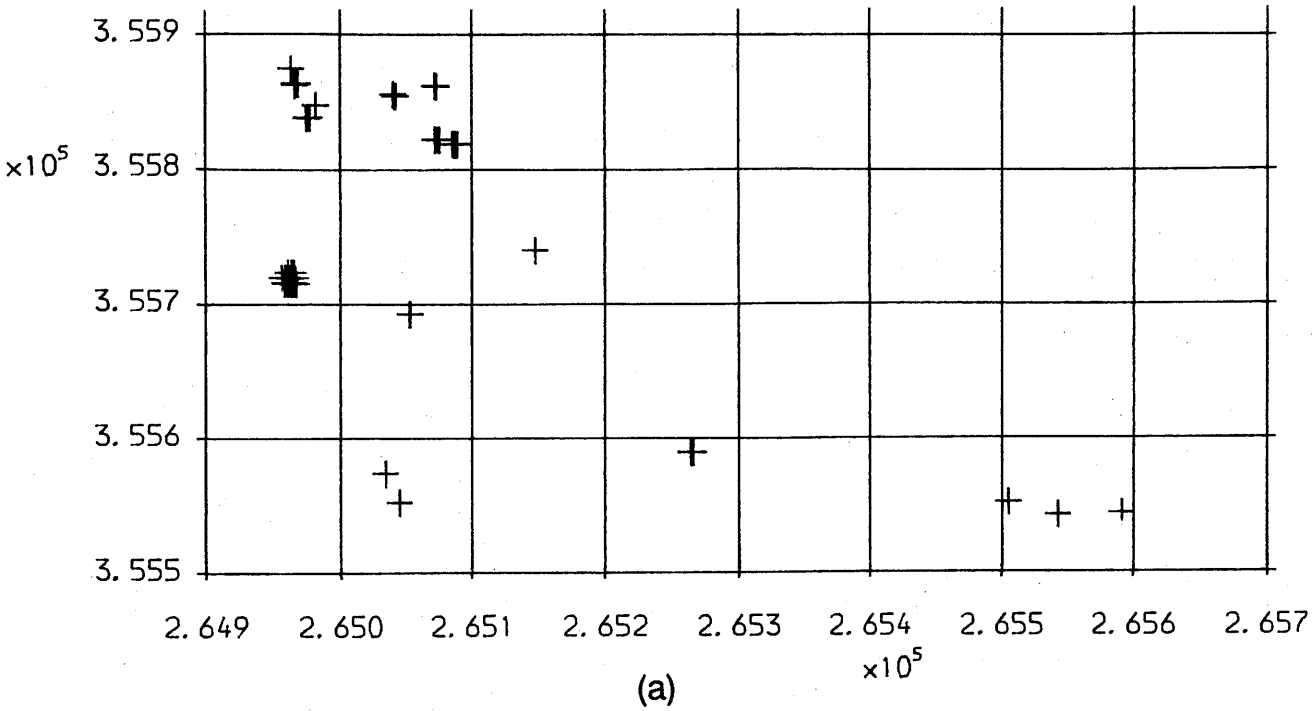


Fig.8.12 Results obtained from the supplementary use of these two algorithms

(a). Detected gross errors;

(b). Contour plot of the data set after gross error removal



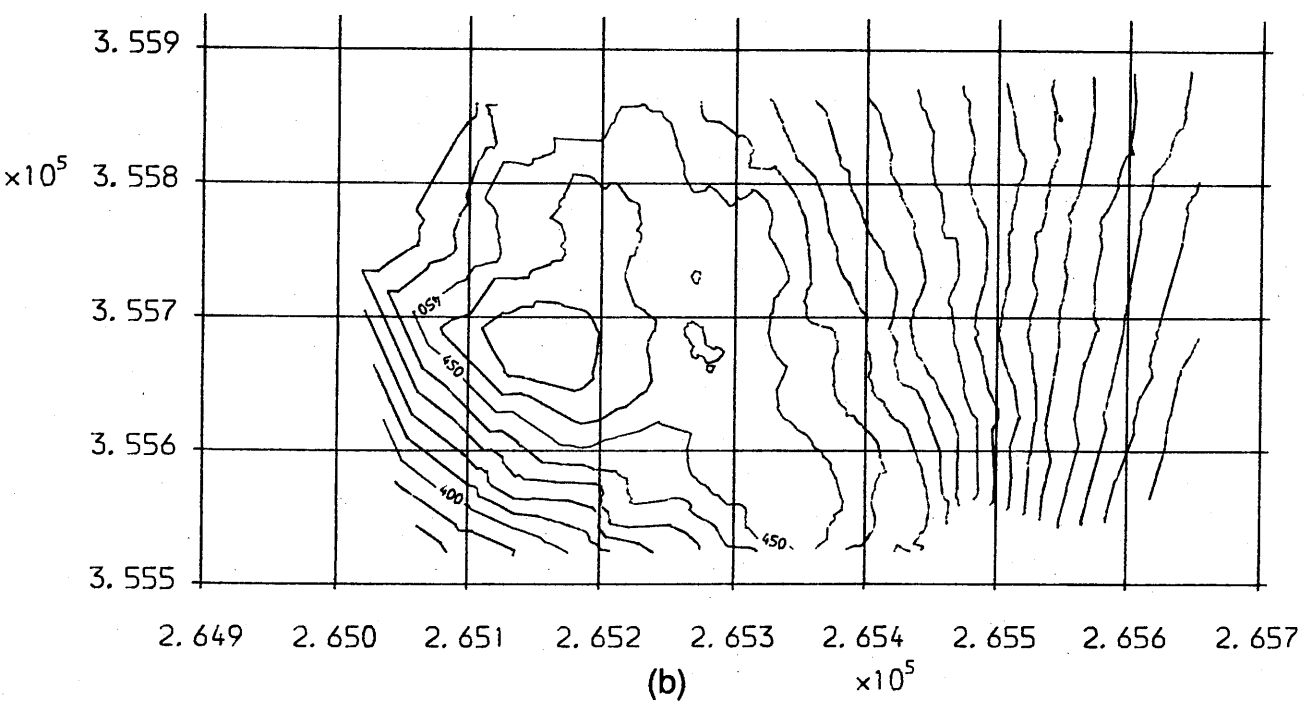


Fig.8.12 Results obtained from the supplementary use of these two algorithms

(a). Detected gross errors;

(b). Contour plot of the data set after gross error removal

8.5.8 Discussion of the computational effort required for detecting gross errors in irregular data sets

Another important question which should be answered here is how much influence is the large amount of cpu time required for detecting gross errors using these algorithms for a given area or for a given number of data points going to have on the users' willingness to adopt these methods. In practice, the required cpu time may vary with many factors such as the distribution of the data points in the data set, the size of the area, the distribution of the gross errors, the initial values given in the program, the effect of increasing the window size, etc. For example, the test results using Algorithm 1 show that the cpu times spent on the two test areas are quite different although the number of total points in both cases are almost the same, due almost entirely to the differences in data density and distribution.

If the data is evenly distributed, then the average value of the area size per point will provide a relatively reliable basis on which to determine the window size, thus saving time in the search for the neighbouring points. If the area is very large and the amount of data is huge, then it will take a long time to search for the neighbours from the whole data set. In this case, if the data is divided into a number of patches, then much less cpu time may be needed. If the gross errors occur in clusters, then a much larger window is needed, which needs much more time, first of all for the point search and then for the subsequent processing. If the initial size of the window area is too small, then a number of iterations will be required before it reaches a reasonable size. This will of course be dependent on the rate at which the size of the window is enlarged - a matter which is only going to be resolved with experience. Anyway, it is a complex matter. Nevertheless, some estimates will be given below although those figures may not be too reliable.

(1). CPU times per point varying with the size of an area

In order to provide more information about the computational side, further tests have been carried out. The purpose of these tests is to provide some information about the relationship between the area size and the cpu time required to check all the points within the area. Therefore, the second test area has been sub-divided into a few patches (which partly overlap on one another). The information about the size of an area and the number of data points within each area is listed in Table 8.7.

The second version of the algorithm was used for these tests. The minimum number of points in a window is specified as 40. The initial area for the window in terms of the half-window size is 3 times the

average value of area size per point. The results are given in Table 8.7. Fig.8.13 shows the relationship between the cpu time per point and the number of points existing within an area.

Table 8.7 Required cpu time and the size of the area

	Area 1	Area 2	Area 3	Area 4	Area 5
Size on Ground	200x400m	200x400m	300x400m	400x400m	500x400m
Size on Photo	1.1x2.2cm	1.1x2.2cm	1.7x2.2cm	2.2x2.2cm	2.8x2.2cm
No. of Points	527	1,121	2,143	3,380	4,188
Total CPU Time	11,886ms	45,597ms	134,515ms	251,480ms	439,997ms
CPU Time/Pt	22.6ms	40.7ms	62.8ms	74.4ms	105.1ms

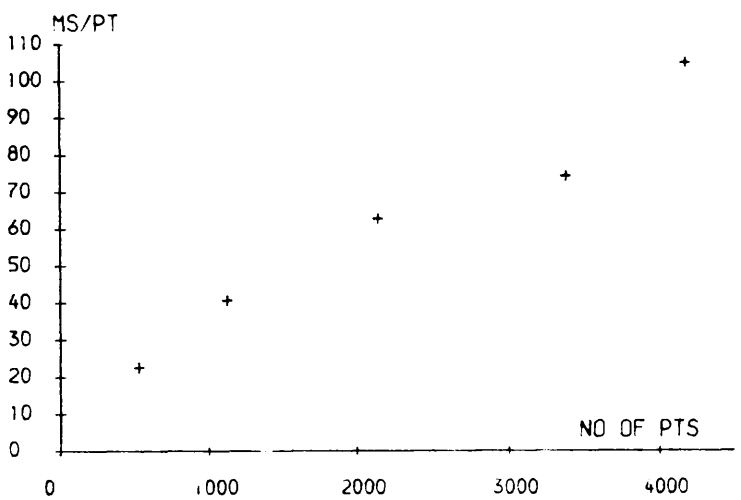


Fig.8.13 The relationship between the CPU time per point and the number of points lying within the neighbouring search area

From this table, it can be seen clearly that the cpu time required to check a point varies with the size of the area within which the search for the neighbours has to be undertaken. Therefore, it is very desirable to use small patches. In this case, as stated previously, a considerable amount of cpu time can be saved. If the sub-area - Area 5 in Table 8.7 - is divided into four patches, each with 1,047 points, then the time required for each patch is 45,597ms. For the four patches, the total time required is 4x45,591ms = 182,388ms. This is much less than the 439,997ms which is required if the large patch size (Area 5 as a single patch) is used. In percentage terms, it amounts to only about 40% of that required to process Area 5 as a single patch. However, even with this substantial reduction, it goes virtually

without saying that the times required for processing are still very large and therefore very expensive to implement.

(2). CPU times required to check all data points of entire model

From the discussion above, it will be seen that the cpu time required for checking all the data points in an entire model will vary considerably with the total number of points and with the size of the patches into which the entire model has been divided. In this context, no prediction about the cpu time required for the entire area of the stereo-model will be wholly reliable. However, as noticed above, it must be kept in mind that the computational time will be very heavy. In any case, the values for the cpu time per point listed in Table 8.7 may help in carrying out some very rough estimates. The following is such an estimate.

Suppose that the data in the second test area which was kindly provided by Mr. Azizi is a typical example. In which case, the density and distribution of the data set for the entire model will be more or less the same. Then extrapolation may be made based on the information in Table 8.7. The largest area listed in Table 8.7 is 2.8x2.2cm. This represents about 1/30 of the entire model area. Therefore, the number of points likely to be generated for the entire model will be about 30x4,188 = 125,640 (points). The cpu times required to check this number of data points using different patch sizes (with a varying number of points in the patch) are given in Table 8.8.

Table. 8.8 CPU times required to detect gross error in the entire stereo-model

Size of patch	527pts	1,121pts	2,143pts	3,380pts	4,188pts
CPU Time/Pt	22.6ms	40.7ms	62.8ms	74.4ms	105.1ms
Total CPU Time	2,993secs =0.83hrs	5,101secs =1.42hrs	7,890secs =2.19hrs	9,347secs =2.60hrs	13,205secs =3.67hrs

8.6 Concluding remarks

In this chapter, some discussions concerning the quality of DTM source data have been made and some algorithms or procedures have been presented which are designed to improve the quality of this data in various ways. First of all, the effect of **random error** on DTM data quality has been investigated. Some examples of these effects have been given and a filter-cum-smoothing procedure has been devised and imple-

mented to remove or lessen their effect. Some experimental results have also been presented which confirm the effectiveness of this type of procedure in certain situations and in certain types of terrain.

For **gross error**, two cases have been considered, i.e. those with gridded data and irregularly distributed data respectively. Some algorithms have been developed and presented for gross error detection and removal. Experimental results prove that they work very well - at least with the rather limited amount of data which was available for the testing.

However, it must be emphasized here that a considerable amount of computation time is needed to implement techniques such as those described in this chapter. In particular, the cpu time required to detect gross errors in an irregular data set may indeed be regarded as quite excessive. Therefore, on the basis of the experience gained in this present project, the advice must be not to apply such an algorithm to an entire model. Instead, it is desirable to carry out a visual inspection to find the general area of difficulty, and then to apply the algorithm to the area which has been identified as probably containing gross errors.

In this chapter, all the discussion and testing have been concerned with the quality of the photogrammetric source data. Starting with the next chapter, the problem of making an estimation or assessment of DTM accuracy will be considered. The discussions will be based both on a theoretical analysis and on experimental investigations. A chapter concerning the relationship between DTM accuracy and sampling distance will be the first step in this investigation.

Chapter 9

Variation of the Accuracy of Digital Terrain Models with Sampling Interval

Chapter Nine

Variation of the Accuracy of Digital Terrain Models with Sampling Interval

9.1 Introduction and background

The various problems concerning the quality of DTM source data have been discussed in the previous chapter, where an attempt was made to remove the gross errors and even the random errors from the source data using certain procedures or algorithms. From this chapter onward until Chapter 13, the discussion will focus on the possible accuracy of DTMs derived from different data sets which are assumed to be free from gross error.

9.1.1 General introduction

As emphasized in the introductory chapter (One), the assessment of the accuracy of DTMs, derived from a set of data which may be acquired from different data sources using different sampling strategies, is one of the two main issues addressed by this project. It has also been emphasized several times previously in this thesis that the accuracy of a DTM is a function of many factors, but mainly the three attributes of the source data (i.e. their acccuracy, density and distribution); the characteristics of the terrain itself; the characteristics of the DTM surface which is constructed from the source data; and the method used for the construction of the DTM surface. The assessment of DTM accuracy made in this project attempts to find the relationship between the accuracy of DTMs and these six accuracy factors. However, the final expression of such a relationship might either be very implicit (i.e. implied) or very explicit (i.e. plainly stated).

The assessment of DTM accuracy can also be carried out either by theoretical analysis, or by experimental investigation or through a combination of both. This chapter and the following chapters will describe the experimental results while the theoretical analysis of DTM accuracy based partly on the results of the various assessments will be discussed later in Chapter 12.

9.1.2 Alternative plans for experimental investigation

Attempting to investigate experimentally the variation of DTM accuracy with any one of the six main factors given above, it is necessary to isolate the effects of each of the other (five) factors on the DTM accuracy by keeping these factors unchanged. In doing so, six possible plans are possible as follows:-

i). The **accuracy of the source data** could be varied while all the other factors remain unchanged. This could be achieved by using aerial photographs with different scales and different flying heights taken over the same test area, employing the same sampling strategy and utilizing the same modelling method to construct the same type of DTM surface.

ii). The **density of the source data** could be varied while all the other factors remain unchanged. This can be achieved by using different sampling intervals in the case of regular grid sampling or composite sampling or alternatively by generating new grids from the original (grid-based) source data.

iii). The **distribution of the source data** could be varied while all the other factors remain unchanged. This can be achieved by using different data patterns such as square grids, rectangular grids, triangular grids, strings obtained from profiling, etc.

iv). The **test area** could be varied but all other factors remain unchanged. This can be achieved very easily by using many different test areas.

v). The characteristics of the constructed **DTM surface** might be varied (e.g. the continuity and smoothness could be changed) while all the other factors remain unchanged. This can be achieved by using different modelling approaches to construct different types of DTM surface as discussed in Chapter Seven.

vi). The **method used for the construction of the DTM surface** could be varied while all the other factors remain unchanged. This can be achieved by employing as many methods as possible, e.g. to construct the DTM surface either directly from the measured data or from a data set which has undergone a random-to-grid processing.

In this project, no attempt has been made to make use of the fifth and the sixth possibilities due to the complexities inherent in the nature of the surface modelling itself. Therefore, a particular type of surface - in this case a continuous surface - was selected for the modelling and this surface has been directly constructed from the measured source data (i.e. without undergoing any random-to-grid interpolation). The question as to why a continuous surface has been selected for the studies in this project has already been discussed and established in Chapter 7. Investigations have therefore been concentrated on the first four plans.

Arising from purely practical considerations of expense and the availa-

bility of data, inevitably some of the tests are not as well designed as they should be. The various sampling **patterns** include square grids, square grids plus strings of feature-specific lines, parallel strings measured by profiling and data strings of constant elevation generated by contours. The **test areas** include three in West Europe and one in the Sudan. The data sets have been acquired from many different sources, e.g. photogrammetric data acquired from aerial photographs with scales varying from 1:4,000 to 1:30,000 and space photographs at 1:950,000 scale obtained using the Metric Camera. Thus the **accuracy** of the source data will also vary considerably. Furthermore, different data **densities** have also been utilized with the acquired source data from different areas.

However, it should be emphasized that the second of the various plans described above is the one which will be given most serious attention and consideration in this project, since, as has been emphasized previously several times, sampling interval is the factor which potentially has a most significant effect on the accuracy of a DTM. A fairly systematic investigation into the variation of DTM accuracy with sampling interval for certain types of terrain using certain data patterns will be presented in this chapter. The remaining tests will be presented later in Chapters 10 and 13.

9.1.3 Basic methodology used for the experimental investigation

For the experimental investigations carried out in this project, the so-called **black box approach** will be used as the basic methodolody of the study. As the name implies, this is an approach which does not take account of the internal composition and workings of the system, which may be unknown, known only partially, or known but ignored. The actual interest of this approach is focused on the input-output behaviour of the system, which accepts one or more inputs and produces one or more outputs.

In the context of the experimental investigation into DTM accuracy, the modelling package is the system (black box), of which much knowledge such as the basic approach for network formation, the possible options of surface types, the subroutines within the programs, the detailed data arrangement within the programs, the variables used in the programs, etc may all be known. However, not all of this information is of interest for a particular project. For example, in this project, only the information about the network formation and the available types of DTM surface need to be known. All other information will be ignored.

The input to the system should, of course, be the DTM source data and the check point data while the desirable outputs in the case of

accuracy testing are the accuracy figures. However, since a DTM system is not usually built specifically for such a purpose but for production purposes, a supplementary program needs to be used to compute accuracy figures from the output of the DTM package (the 3-D coordinates of the DTM). This program can also be considered as an integrated part of this system.

In this chapter, information about the source data acquired from photography taken at different scales for three different test areas will be given first. Then a discussion about how to generate new grids from the original grids will be presented. Next, the DTM package and the program for the computation of accuracy figures (comprising the black box) used for this study will be briefly described; after that, the output from the black box - the accuracy figures - will be presented. Finally, an analysis will be conducted to try to establish the inter-relationship between the input and the output of the black box (the DTM system).

9.2 Description of test data

The source data sets used in this experimental test are some of those which were used for the ISPRS DTM test (Torlegard et al, 1986) although all the processing and the tests based on these data sets were carried out independently by the author.

9.2.1 Test areas

In the ISPRS DTM test which was conducted by Working Group 3 of the Commision III, six areas were used. Some data sets for five of these six areas were made available to the author for use in this project. However, it was found that only three of these data sets, i.e. those for the Uppland, Sohnstetten and Spitze areas, were suitable for the studies to be described in this chapter due to problems related to data patterns which had been collected and/or certain limitations of the DTM package used in the tests.

The descriptions of these three test areas have been given by Torlegard et al (1986). However, for convenience, a brief summary is included here in the form of Table 9.1.

Contour plots and isometric views of the DTM surfaces for each of these three areas are given in Figs.9.1, 9.2 and 9.3. The data sets and the corresponding contour plots had been measured on a Zeiss Oberkochen Planicomp C-100 analytical stereo-plotter at the Technical University of Munich and were made available to the author through the courtesy of Prof. H. Ebner and Dr. W. Reinhardt.

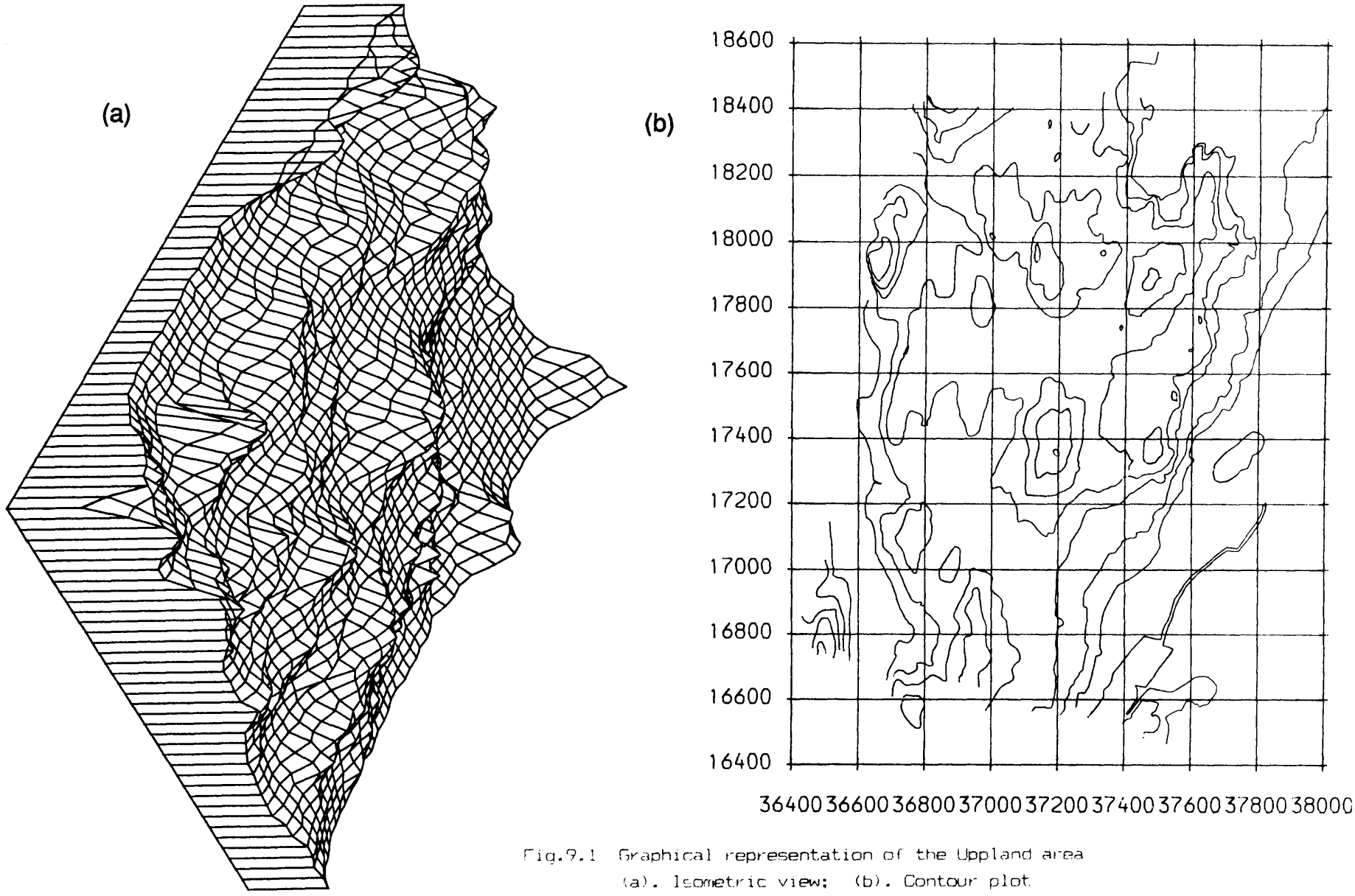


Fig.9.1 Graphical representation of the Uppland area
(a). Isometric view; (b). Contour plot

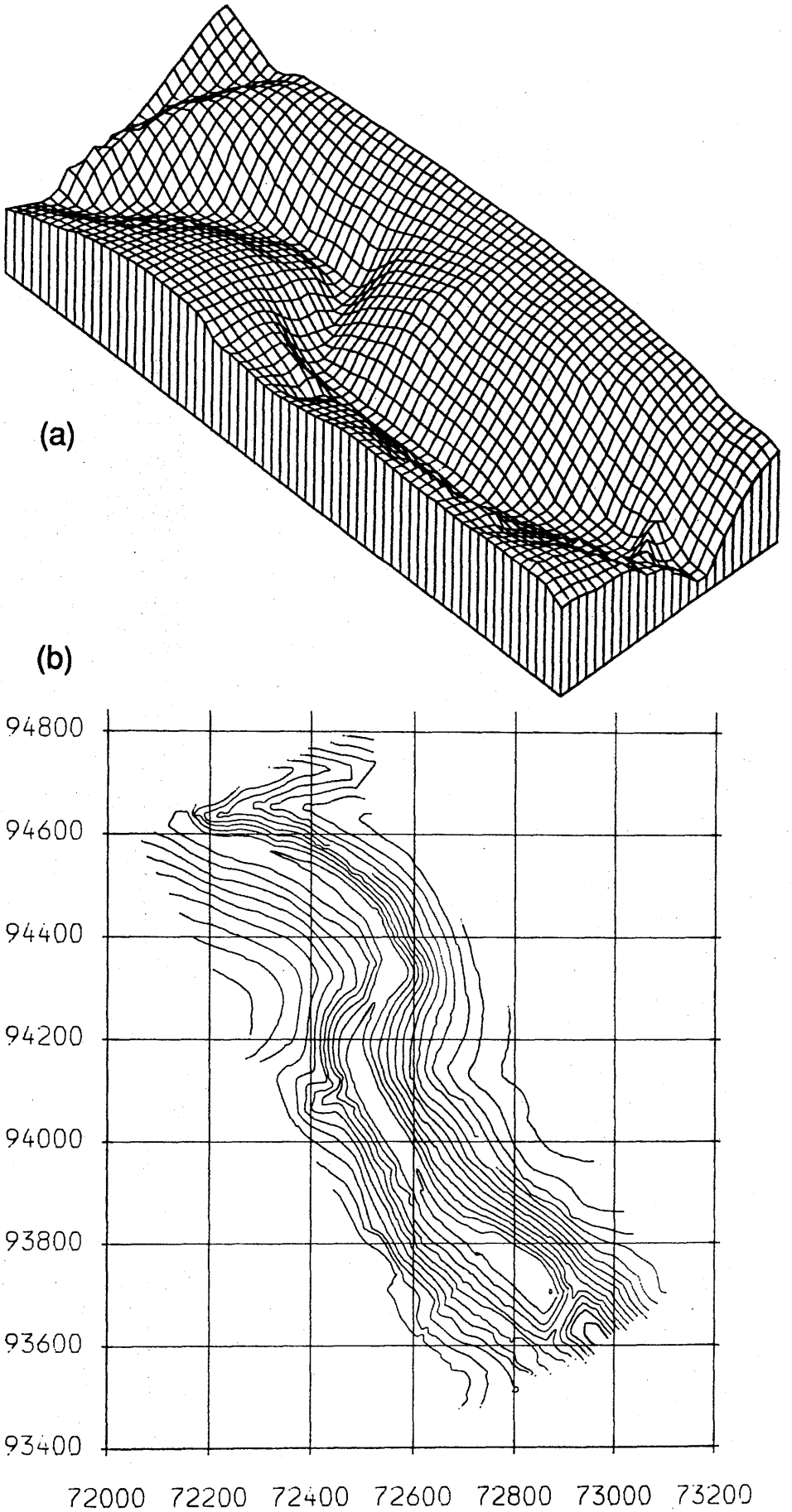


Fig.9.2 Graphical representation of the Sohnstetten area
(a). Isometric view; (b). Contour plot

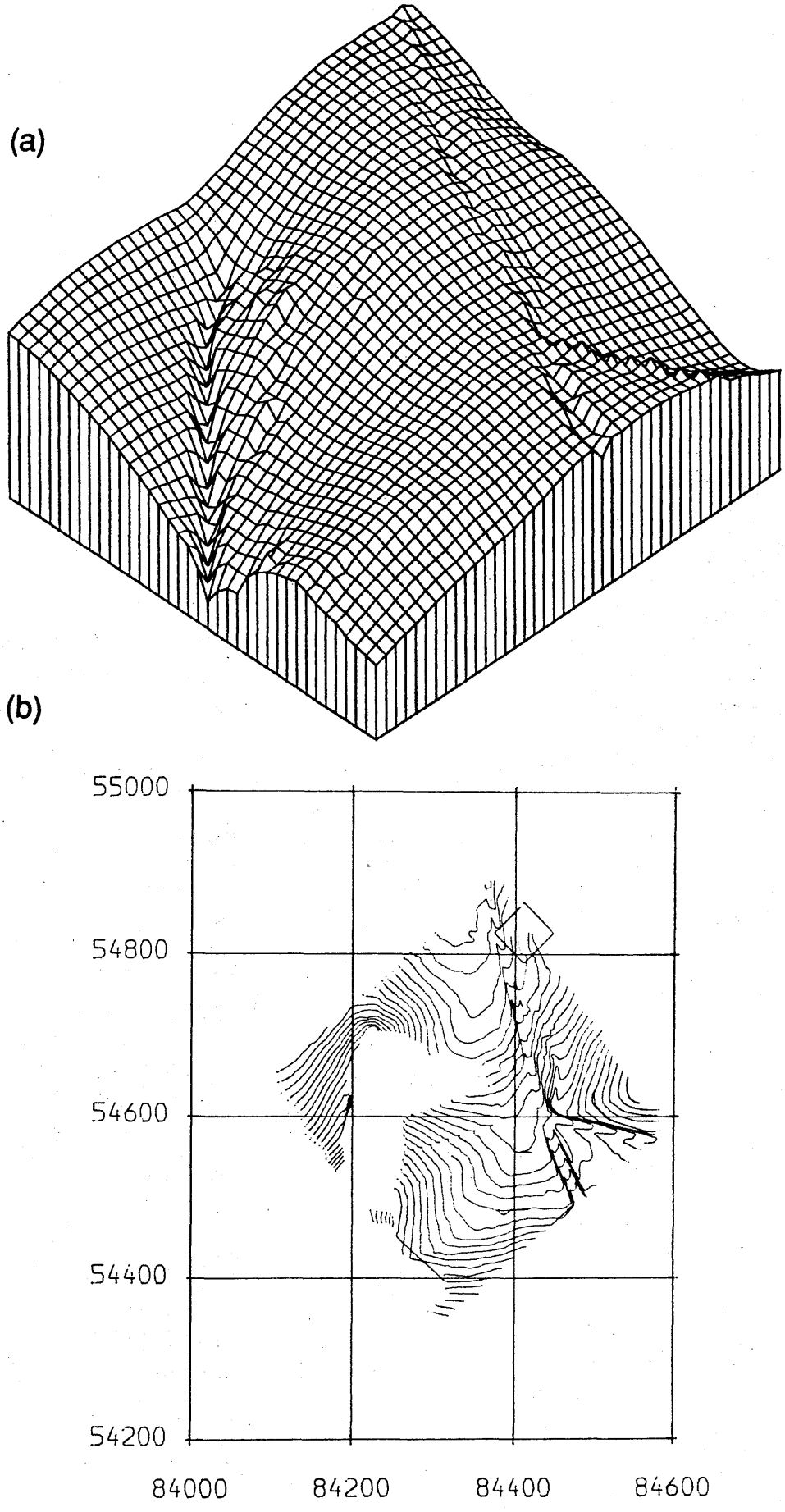


Fig.9.3 Graphical representation of the Spitze area
(a). Isometric view; (b). Contour plot

Table 9.1 Locations and descriptions of the test areas

Test Area	Description	Height Range
A. Uppland (Sweden)	Farmland and forest	7 to 53 m
E. Sohnstetten (W.G)	Hills of moderate height	538 to 647 m
F. Spitze (W.Germany)	Smooth terrain	202 to 242 m

9.2.2 Source data sets

The data sets had been acquired on the Planicomp C-100 using composite sampling. The gridded data sets were stored separately from those of the feature-specific points (including the points measured along break lines and form lines). Figs.9.4 to 9.6 show the distribution of the data points for each of the three test areas. The grids were more or less arbitrarily orientated. For the area of Uppland only, two grids were measured with origins shifted against each other by 20m in the X and Y directions; however, the data points shown in Fig.9.4(a) belong to only one of these two grids. The information about the grid sizes, together with the photo scales, is given in Table 9.2.

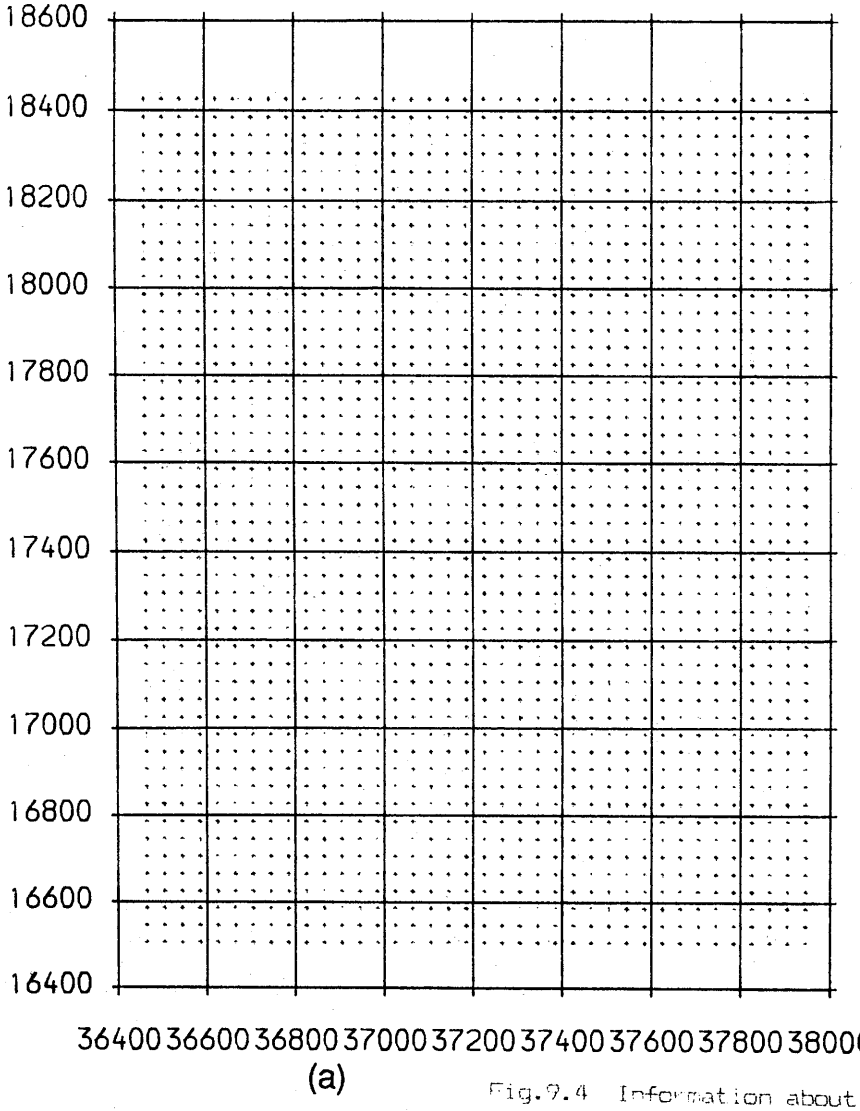
Table 9.2 Data sets used for testing

Test Area	Location	Photo Scale	H	Grid Interval
A	Uppland (Sweden)	1: 30,000	4,500m	40m (2)
E	Sohnstetten (W.G)	1: 10,000	1,500m	20m
F	Spitze (Germany)	1: 4,000	600m	10m

Table 9.3 Accuracy of source (raw) data

Test Area	Location	Orientation accuracy		Data accuracy
		Planimetric	Height	Height
A	Uppland	±0.474m	±0.497m	±0.67m
E	Sohnstetten	±0.110m	±0.057m	±0.16m
F	Spitze	±0.072m	±0.048m	±0.08m

The standard errors (standard deviations) of unit weight for the absolute orientation at ground scale (in which the co-ordinate obser-



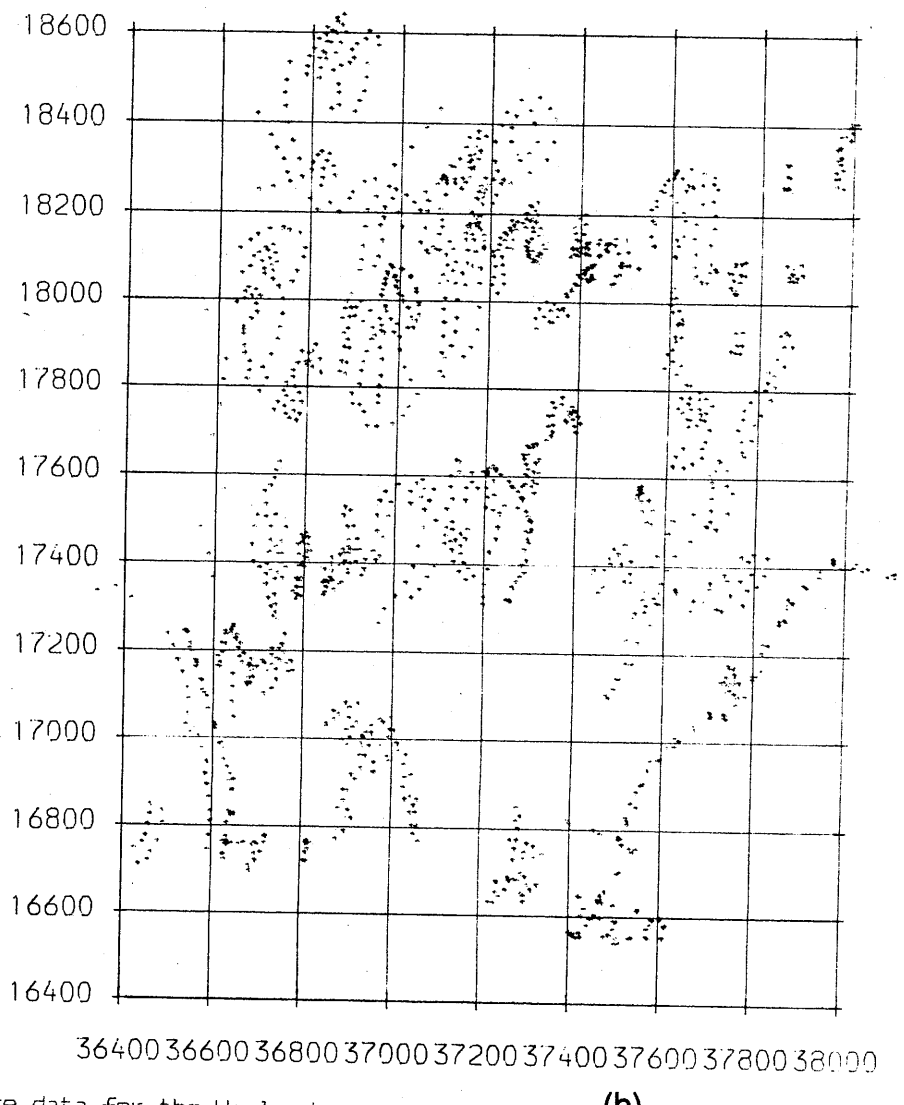
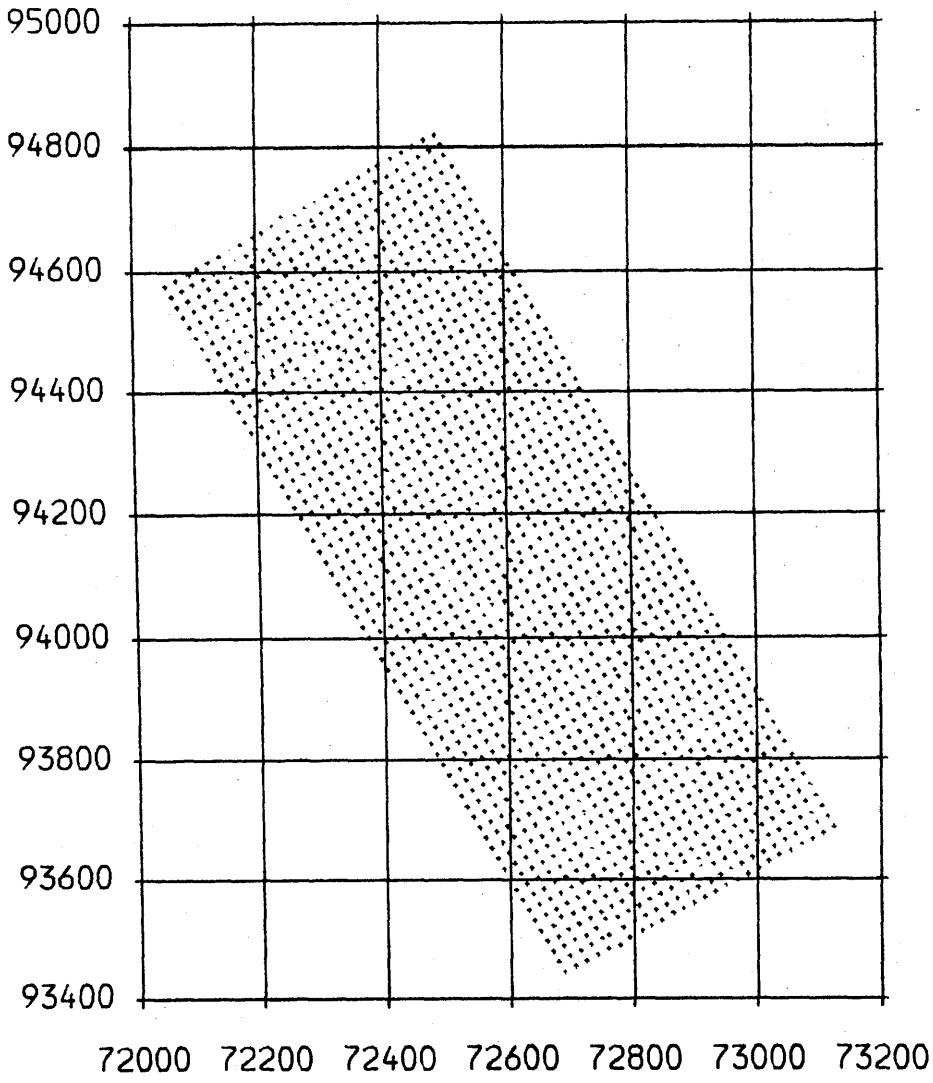


Fig.9.4 Information about the source data for the Uppland area
(a). One of the two 40m grids; (b). Feature-specific points and lines



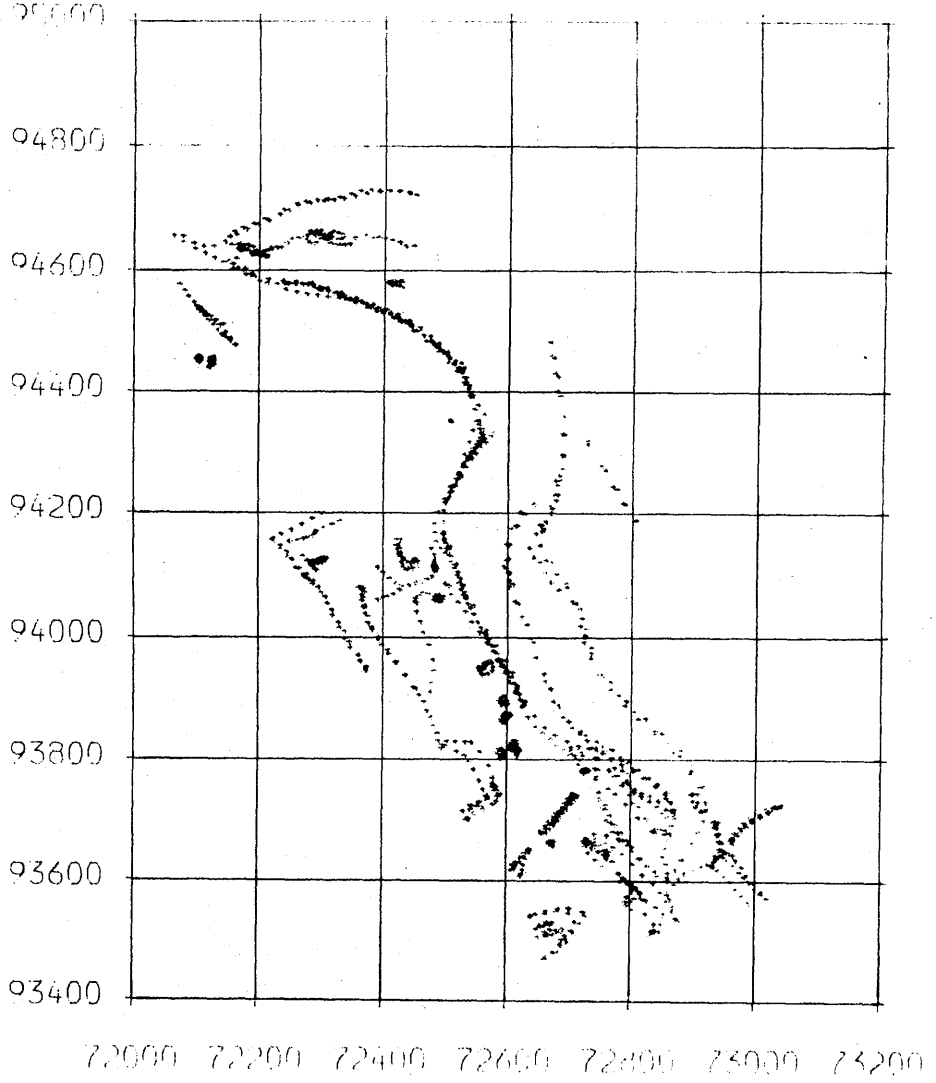


Fig.9.5 Information about the source data for the Sohnstetten area
(a). Gridded data; (b). Feature-specific points and lines

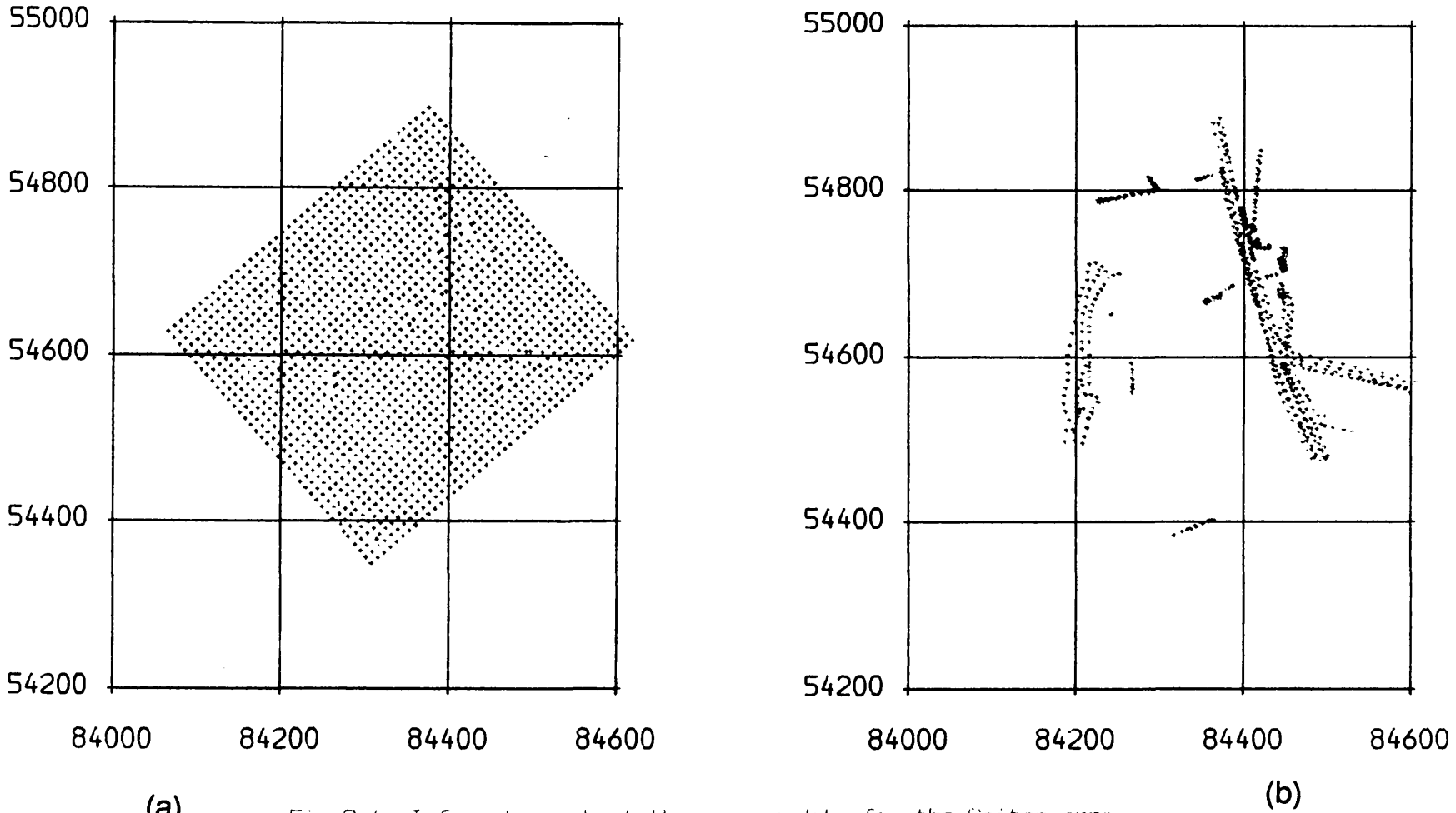


Fig.9.6 Information about the source data for the Spitze area (a). Gridded data; (b). Feature specific points and lines.

vations are given unit weight) are quoted in Table 9.3. The height accuracy of the raw data has been estimated according to the discussion in Chapter 5 and is also given in Table 9.3.

9.2.3 Check points

The check points used in this test were measured from much larger scale photography in the Royal Institute of Technology (the ISPRS DTM test centre) in Stockholm and were kindly made available to the author through the courtesy of Prof. K. Torlegard and Dr. M. Li. These check points were used in the ISPRS tests and are also referred to as "ground truth" in the paper by Tolegard et al (1986).

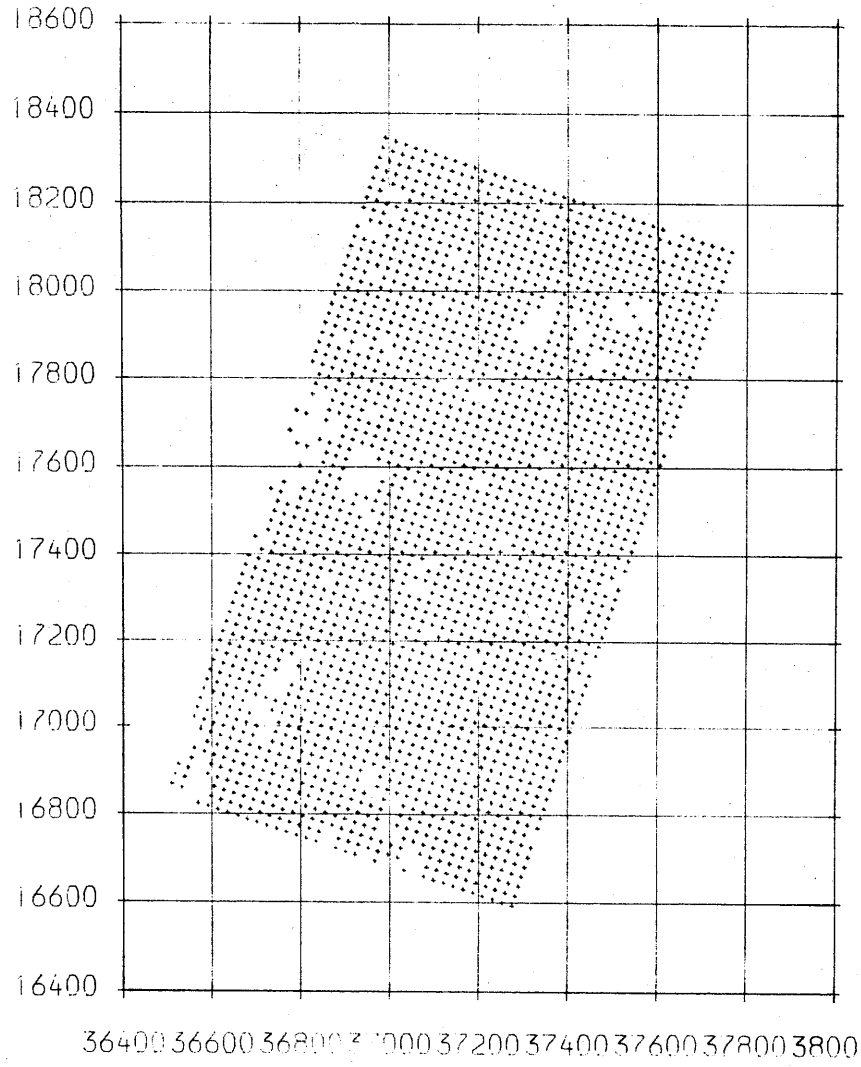
The check points are located in a grid pattern which is orientated in an arbitrary direction. This means that the grid directions of the check points are not parallel to those of the source data or to the X and Y coordinate directions. Actually, not every grid node was measured; those points which are located in the areas where no reliable measurements can be made such as areas covered by bush or trees and those with extremely steep slopes were not measured and their height values were recorded as zero. Figs.9.7(a), 9.8(a) and 9.9(a) show the distribution of the check points for each of the three test areas. In fact, for each test area, more than 1,700 points were measured on analytical plotters in a static mode. Aerial photographs with scales much larger than those used for the acquisition of the DTM source data were used in order to ensure that check points did indeed have a much higher accuracy. Detailed information about the characteristics of these check points has been given by Lindgren (1983). Table 9.4 shows some of this information.

Table 9.4 Information about check points

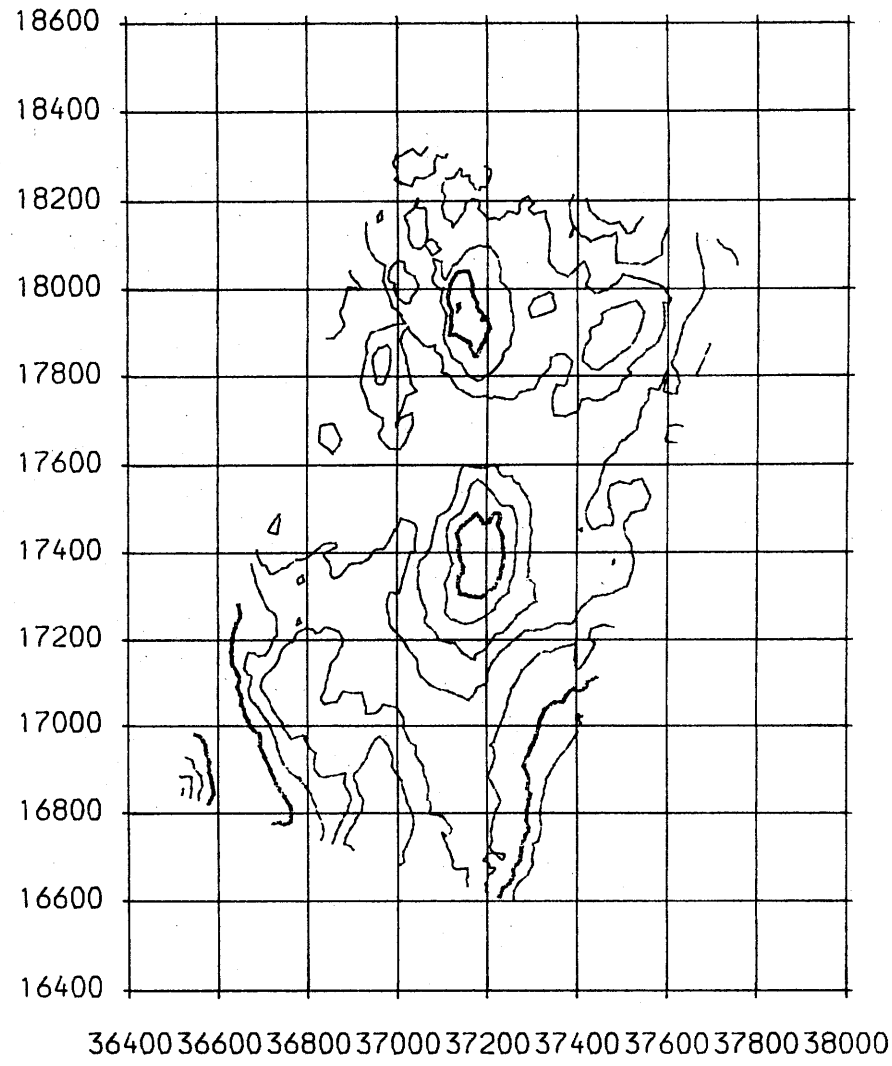
Test Area	Photo Scale	H	No. *	RMSE	Emax
Uppland	1: 6,000	900m	2,314	±0.090m	0.20 m
Sohnstetten	1: 5,000	750m	1,892	±0.054m	0.07 m
Spitze	1: 1,500	230m	2,115	±0.025m	0.05 m

* No. in this table denotes the number of check points.

The accuracy of the check points was checked in the test centre and the results are given in Table 9.4 under the headings of RMSE (root mean square error) and Emax (maximum error). The check points are not distributed over the whole of each test area but only over a part of each area. Figs.9.7(b), 9.8(b) and 9.9(b), which are the contour maps

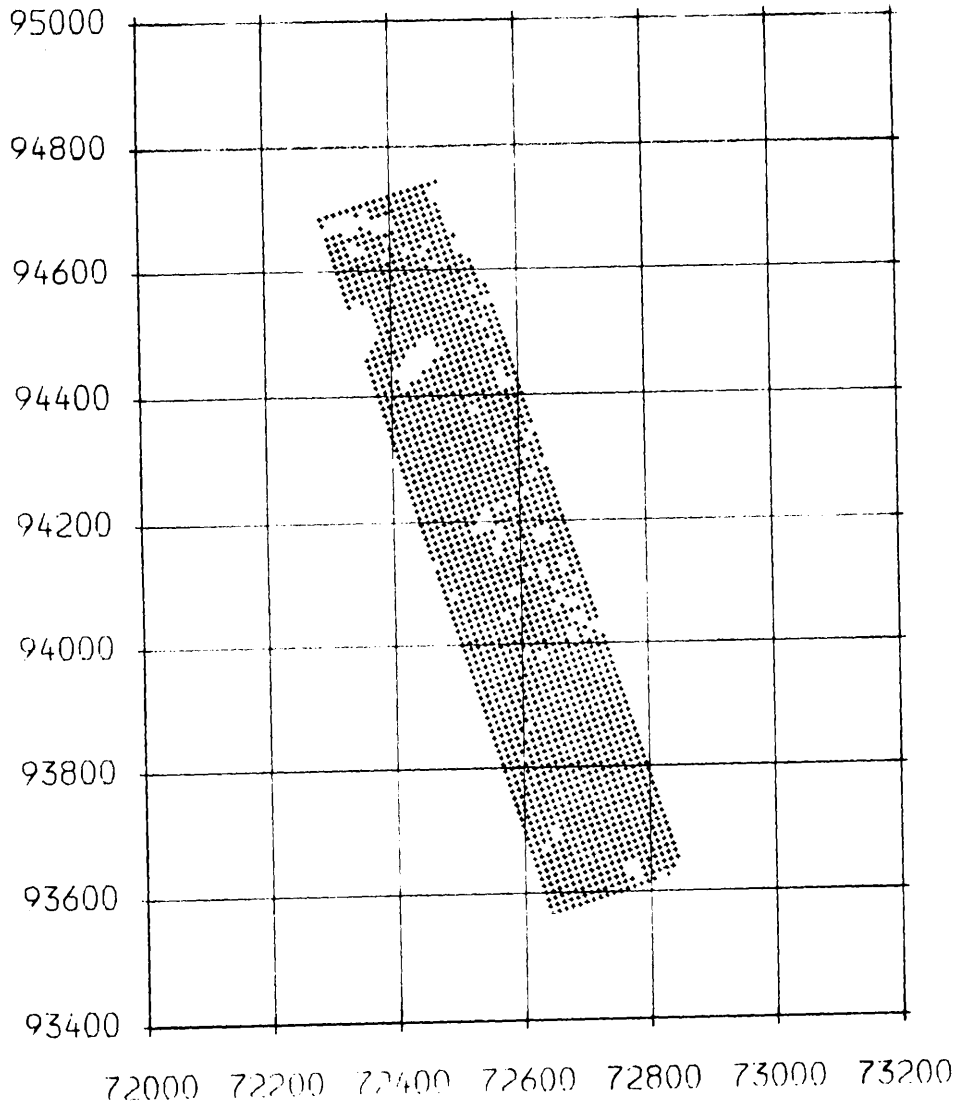


(a)

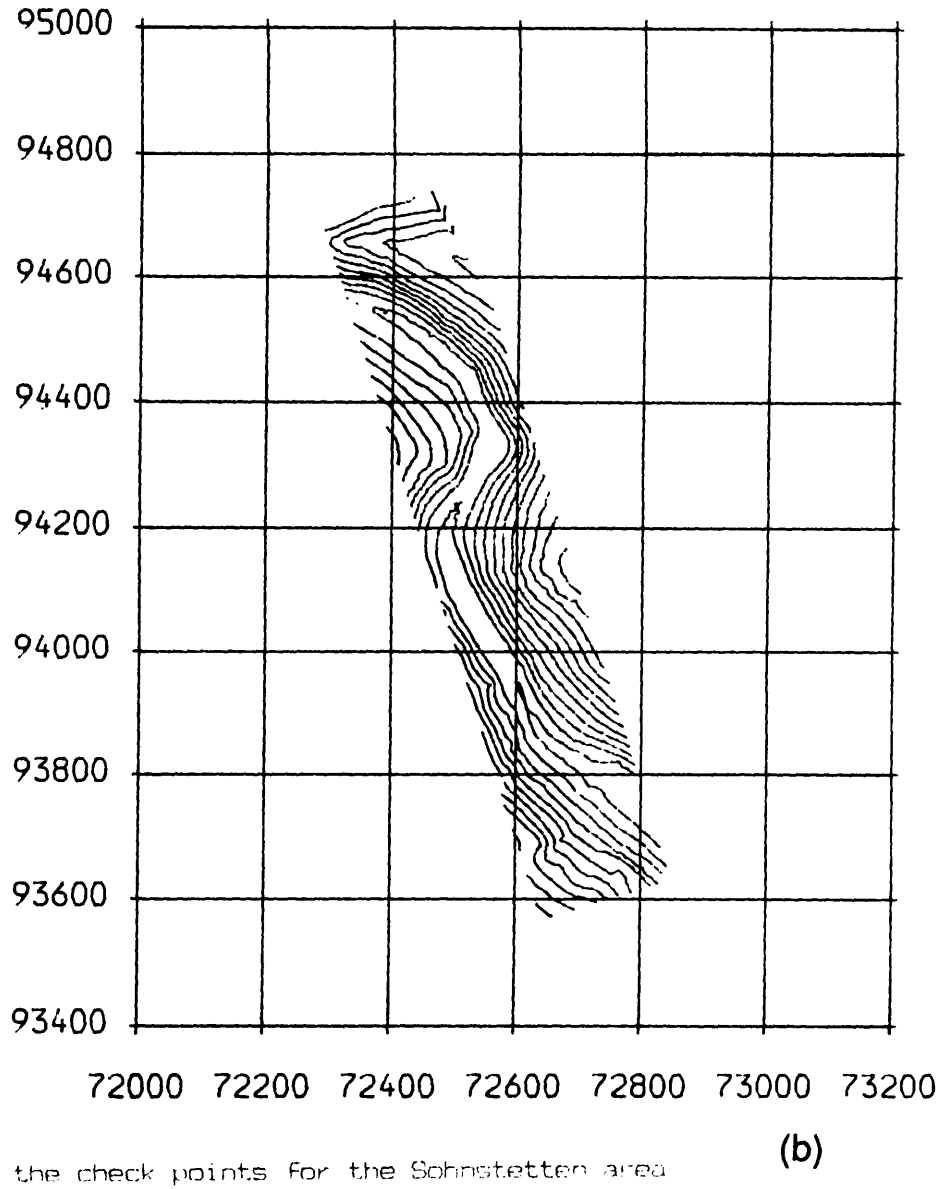


(b)

Fig.9.7 Information about the check points for the Uppland area
(a). The measured gridded points; (b). Contour plot



(a)



(b)

Fig.7.8 Information about the check points for the Sohnstetten area (a). The measured gridded points; (b). Contour plot

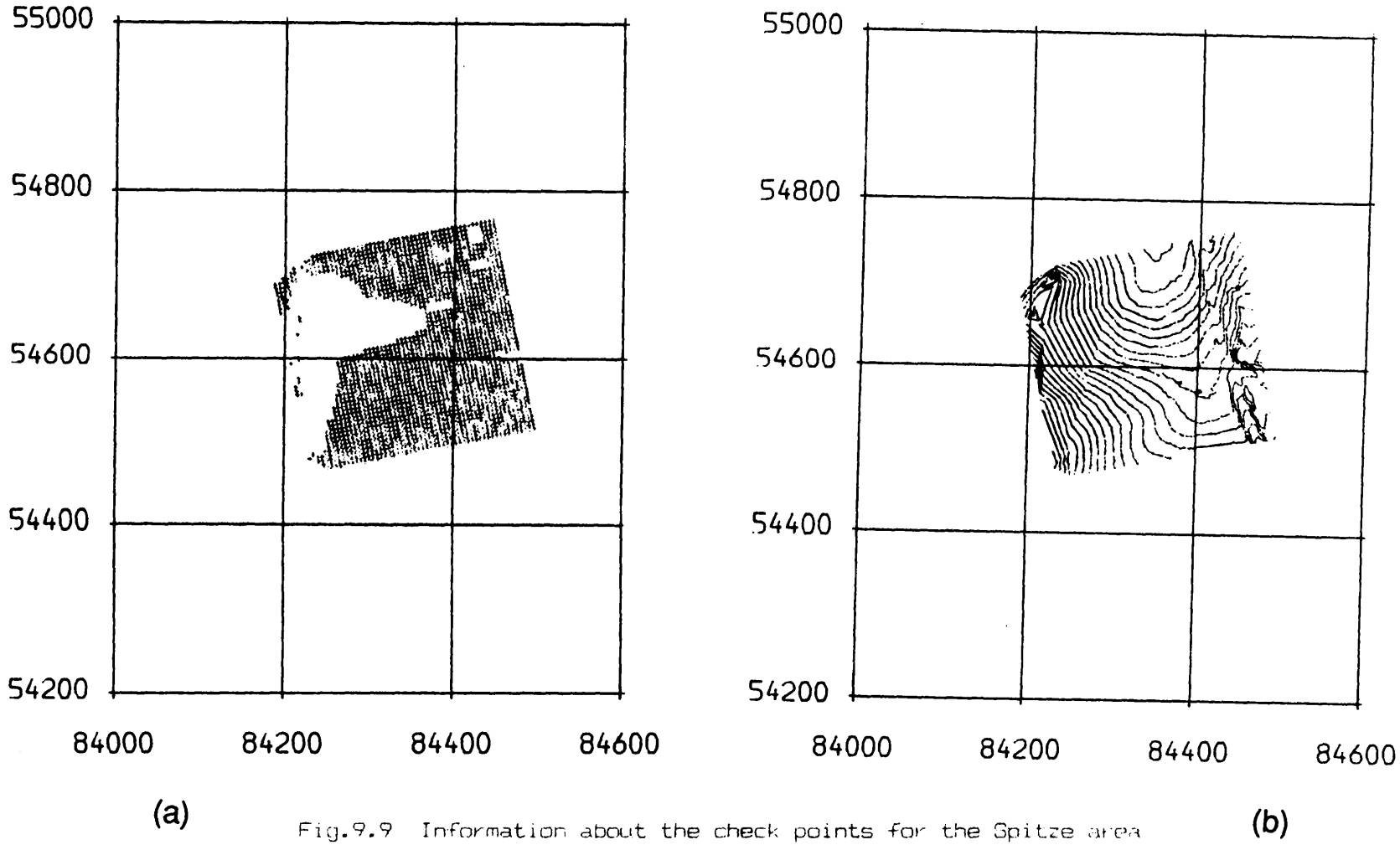


Fig.9.9 Information about the check points for the Spitze area
(a). The measured gridded points; (b). Contour plot

produced from each set of these check points, indicate the variations in the terrain elevations within the areas which are covered by the check points for each test area. Inspection of these contours shows that some segments of these contours are unnatural and artificial in appearance. The presence of such artificial features indicates that no points have been measured in such areas due to woodland or too steep terrain.

9.3 Generation of new grids from the original gridded data sets

As has been stated in the introductory section of this chapter, the purpose of this experimental investigation is to find some relationship between DTM accuracy and the density of the data set for certain types of terrain using certain data patterns. In fact, as described in the previous section, only three types of terrain and two data patterns are available to check this specific case.

The type of terrain is specified by its slope as suggested in Chapter 3. Therefore, representative slope values have been derived from the contours for each of these three areas, with values of 6^{o}, 15^{o} and 7^{o} for the Uppland, Sohnstetten and Spitze areas respectively. The two types of data pattern are, of course, gridded data and composite data. The data density can be defined in terms of sampling interval since the source data comprises either gridded data or gridded data plus strings.

In order to reach the goal of this investigation, it would have been best if a few sets of gridded data (then composite data by adding string information) had been measured for each of three areas. This had not been done - which was not in any way the fault of the participants in the test since it was not a requirement of the ISPRS test. So for each area, only a single specific sampling interval was used. Therefore, a solution had to be found to overcome this problem associated with this specific goal of this investigation.

There are possible two solutions. The first is to measure **more data sets** for each test area using the same diapositives on a photogrammetric instrument. The second is to generate gridded data sets with **different grid intervals** by changing the sampling pattern of the original measured data using wider intervals or other selection criteria. Of course, the latter is not only a quicker and more economic solution but also a more appropriate solution since, in this way, the accuracies of the (new) generated gridded data will be the same as those of the original gridded data. Therefore, the latter procedure has been selected for this study.

The method which has been implemented to generate gridded data sets

with different grid intervals is illustrated in Fig.9.10. The first four figures, Figs.9.10(a) to (d), show the generation of new grids with twice the interval of the original grid. In these diagrams, those points marked with "o" are those which have been retained by selecting every other point along both the row and column directions. They also show that, by using a different starting point, four such alternative data sets can be generated from the original gridded data.

Figs.9.11(e) and (f) illustrate the generation of grids with a grid interval of 1.414 (the square root of 2) times the original one. Again, the symbol "o" denotes those points which have been selected. The directions of the new grids have been rotated and orientated along the diagonal directions of the original grids. It can be seen from these diagrams that two such new grid data sets can be generated from the original gridded data. Actually, this method can also be viewed as producing a regular triangular network comprising isosceles right-angled triangles with each hypotenuse lying along the row direction. It can also be viewed that a regular grid cell is constructed by two such triangles with their common hypotenuse as a diagonal of the grid cell.

This selection can also be considered as adding together two grids with a grid interval of twice the original one and with the origins shifted against each other by the original grid interval. Therefore, very similarly, for the Uppland area, a new grid with a grid interval of about 28.28m (0.707x40m) can be formed by adding together the two grids with a grid interval of 40m.

Table 9.5 No. of Data sets generated from the original data sets

Grid Interval	Uppland	Sohnstetten	Spitze
10.00 m	-	-	1 *
14.14 m	-	-	2
20.00 m	-	1 *	4
28.28 m	1	2	-
40.00 m	2 *	4	-
56.56 m	4	8	-
80.00 m	8	-	-

Where * denotes that the set is the original one.

In practice, the measurement process has been carried out twice in a forward and in a backward direction. Therefore, in the generation of

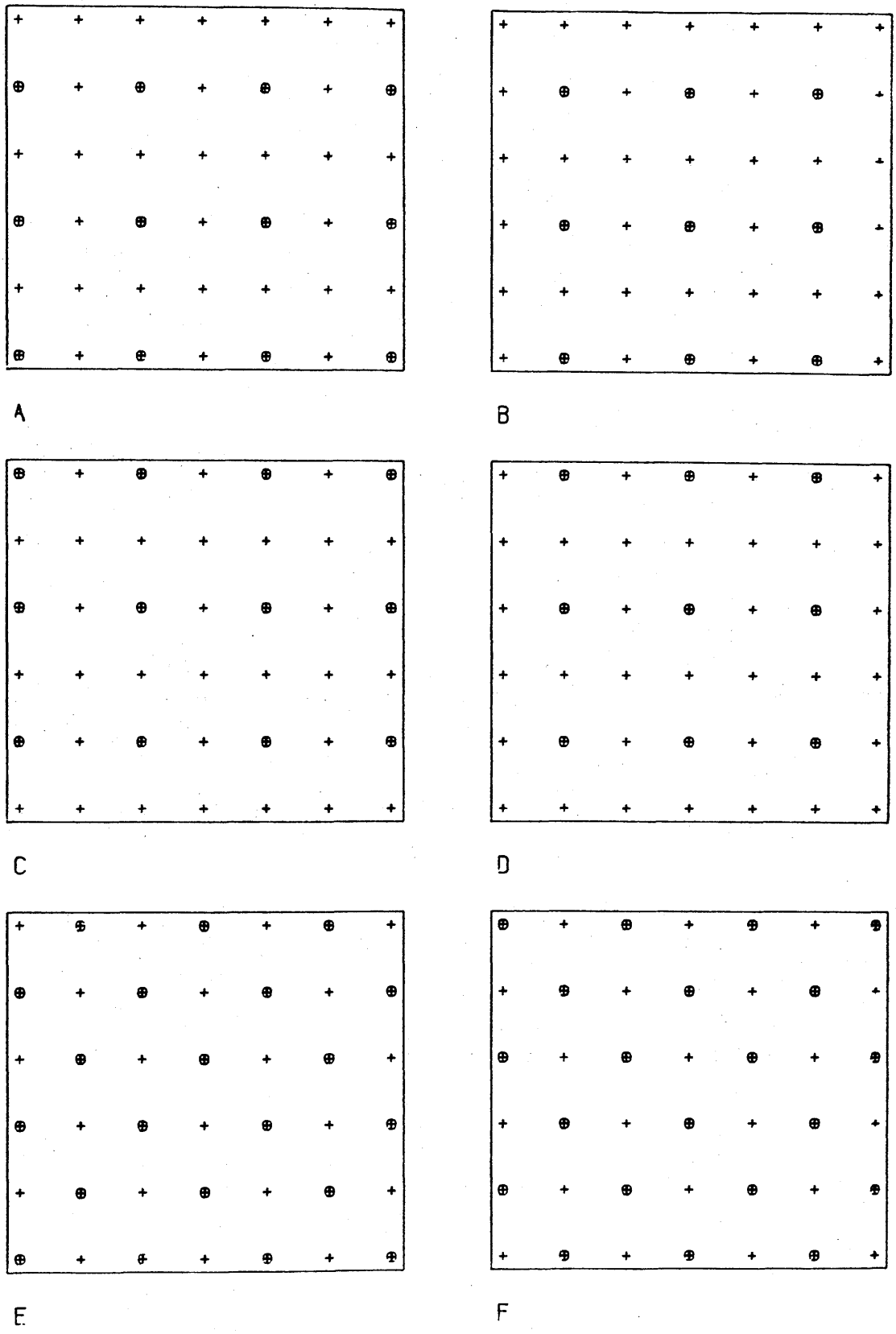


FIG. 9.10 GENERATION OF NEW GRIDS WITH DIFFERENT INTERVALS

a - d : new grids with doubled grid interval

e & f : new grids with $\sqrt{2}$ times greater grid interval

new grids, a corresponding rearrangement had also to be made.

Table 9.5 lists all the test data sets generated from the original data sets (including the original data sets themselves), which are to be used in this experiment.

9.4 Some features of the modelling system - the PANACEA package

In the previous section, the input to the black box has been described. In this section, a brief discussion of some features of the black box itself - the modelling package - will be given.

9.4.1 General structure of the PANACEA package

The modelling package used for the studies of this project is called PANACEA, which is a triangulation-based DTM package. The PANACEA system was developed by McCullagh (1983) of the University of Nottingham. The first version of this package was implemented some years ago in the ICL 2988 mainframe computer at the Univerisity of Glasgow and transferred to the new ICL 3980 machine when it was installed two years ago. The individual programs of this package implemented on the ICL mainframe machine call routines from GINO (a general graphics system) for graphics realisation, either on a graphics terminal or on a hard-copy plotter. The general process of modelling is as follows:

That is, the input data points are, first of all, triangulated by a program called PANIC. Then DTM grids are interpolated from the surface constructed from the triangular network by the program PANDORA. Finally, any desired product, such as an isometric view and/or a contour line plot, will be produced from the surface constructed from the (interpolated) gridded data, in this case, by the corresponding programs - PANORAMA and PANACHE respectively. In addition, a graphics editor called PANDEMON allows users to carry out some editing work, e.g. to display the structure of the triangular network and contouring from the triangular network; to allow the insertion of additional elevation values or strings; and to calculate and display the height of any point lying within a triangular area. Another program called PANEL merges the individual (smaller) grids interpolated by PANDORA into a large single grid. Fig.9.11 shows the inter-relationships between these programs described above.

Actually, for the new version of the PANACEA package, a contouring

program PANTEC has also been developed for the production of contour plots from the DTM surface which is directly constructed from the triangular network formed by PANIC. This program was modified and implemented by the author on the ICL 3980 mainframe machine, on which all the contour plots in this thesis were produced.

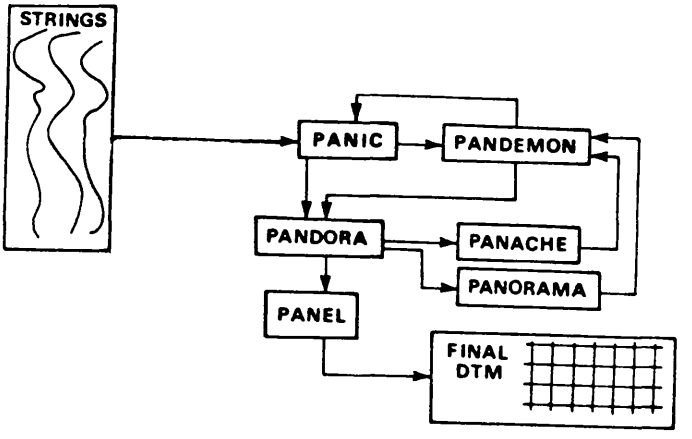


Fig.9.11 The structure of the PANACEA system

9.4.2 Selection of features from the modelling system for this study

As described above, the PANDORA program in the PANACEA system carries out the interpolation of gridded points from the surface constructed from the triangular network. This surface could either be a continuous surface comprising a series of contiguous triangular facets or or a smooth patch surface, given as options. Next the grid intervals need to be defined. The program will ask the user to enter the numbers of rows and columns needed to cover the area of interest, as well as the grid intervals in both the X and Y directions. After these have been defined, the height matrix from the grid data will be supplied as the output. That is to say, PANDORA will only interpolate the data points in the form of grids which are parallel to the X and Y coordinate axes. Obviously, this did not suit the interpolation of DTM points in the case of this project since the check grids in this test were arbitrarily orientated and the check points in the other tests of this project were randomly located.

Another possible option is to use PANDEMON. The interpolation of the elevation value for each specific point again takes place using the linear facets constructed from the triangular network. However, this program only displays the height of a point at the cursor position on the screen by moving and hitting the cursor and not by accepting the input from a file.

Therefore, some modification either to PANDORA or to PANDEMON was necessary for the purpose of interpolating the DTM values which correspond to the check points used in this project. Actually, it was decided to modify the PANDEMON program. After modification, the program accepts a file containing the coordinates of the check points, then computes the new height (Z) values using linear interpolation for each pair of X and Y coordinates. Indeed, the modified PANDEMON program has been used throughout this project to interpolate the DTM data points which correspond to the check points. The output from the PANDEMON is then compared with the input to the program to compute the errors at the DTM points. From these errors, statistical values for overall accuracy have been computed.

9.5 Results of the experimental tests

The characteristics of the input data and the so-called black box itself (the PANACEA system) have been discussed in the previous section. In this section, the output - the accuracy results - from the experimental tests will be presented.

9.5.1 Alternative measures of DTM accuracy

In order to present the results of the experimental tests informatively, some comprehensive measures of DTM accuracy should be used. The accuracy measures which can possibly be used in practice have been described and the usefulness of these measures has also been discussed by the author (Li, 1988a). A brief discussion will also be given here.

Considering the case of the experimental tests, the set of values of the height differences between the points on the DTM surface and those measured on the actual terrain surface are used to compute the required accuracy figures. Therefore, the height differences between the terrain surface and the DTM surface can be considered as a **random variable**, DH, in the context of the accuracy test. The set containing all possible values of the variable is referred to as the **sample space** (or **population**). Therefore, the set of height differences computed through the use of check points is a subset of the population. The **characteristics** of this population are then studied by analyzing the subset sampled from it. In particular, an attempt is made to find:

i) the magnitude of the random variable; and
ii) the spread or dispersion of the random variable.

To measure the former, some parameters can be used such as the **extreme values** (DH_{min} and DH_{max}), the **mode** (the most likely value), the **median**

(the frequency centre) and the **mathematical expectation** (weighted average or **mean**). The two extreme values can be used as such a measure in the sense that they indicate the general location of all the other values, i.e. they lie between these two values. The mode is a measure in the sense that it is the value of the quantities most often observed (i.e. with the largest probability). The median is a value such that observations (the values of DH in this case) above it would occur with the same frequency as those below it. The mathematical expectation is the "balance" point of the distribution of a random variable, like the centre of gravity in physics. In the context of a DTM accuracy test, the simple arithmetic mean (or simply **mean**)of the DH values is the mathematical expectation of random variable DH. The last has been adopted in most cases. However, if the distribution is normal, these parameters have the same value.

To measure dispersion, some parameters such as **range** (DH_{max} - DH_{min}), the expected **absolute deviation**, and the **standard deviation** can be used. Among these, the last is the most useful in most cases.

In summary, in the case of this study, the extreme values, the mean and standard deviation of the test results have been used. In addition, the values of root mean square error have also been computed since this is a traditional measure which is still widely used at present time. The **annotation** is as follows:

RMSE - root mean square error;
Mean - average value of residuals (DH values);
SD - standard deviation of the residuals from the mean;
Emax - one of the two extreme values - the maximum DH value (or residual) of the distribution; and
-Emax - the other extreme value - the minimum DH value of the same distribution.

The values of all these parameters given in the corresponding tables are expressed in metres. The symbol "$\pm$" before the values for SD and RMSE has been omitted simply for convenience in this study.

The **occurrence frequencies** of large residuals have also been computed and are given in the corresponding tables for the reasons that the information about the occurrence frequencies of large residuals can also give investigators some general hint about the accuracy of the DTMs and that this information will be useful for the theoretical analysis of DTM accuracy which will be conducted later in Chapter 12.

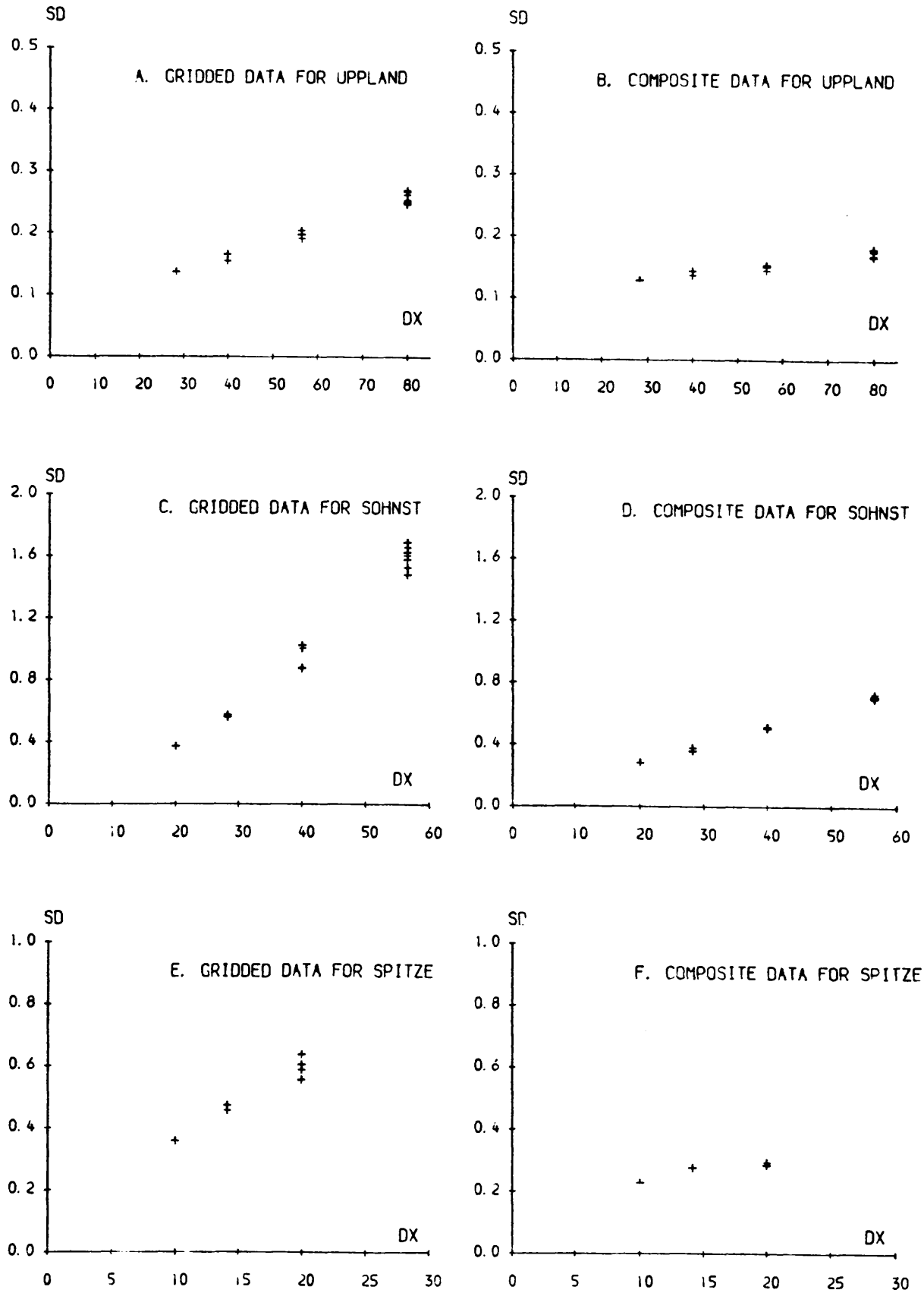


FIG. 9.12 VARIATION OF DTM ACCURACY WITH SAMPLING INTERVAL

UNIT IS PER MIL OF FLYING HEIGHT

9.5.2 Accuracy of the DTM derived from regularly gridded data sets

The accuracy results of the digital terrain models formed from the regularly gridded data sets from the ISPRS experiment are given in Tables.9.6 and 9.7. Diagrammatic representations of the test results are presented in Figs.9.12(a), (c) and (e), while the occurrence frequencies of large residuals are given in Tables 9.8 and 9.9.

Table 9.6Accuracy results from gridded data sets for the Uppland area

Grid No.	Data Set		RMSE	SD	MEAN	+Emax	-Emax
40m G1+G2	28.28m Grid		0.69	0.63	0.26	3.26	-5.78
40m Grid 1	40m Grid		0.77	0.76	0.10	3.49	-6.42
	56.56m G	1	0.91	0.91	0.06	3.41	-6.42
		2	0.94	0.94	0.08	3.87	-6.51
	80m Grid	F1	1.24	1.24	0.07	6.49	-6.83
		F2	1.15	1.15	0.05	4.68	-7.18
		B1	1.14	1.14	0.06	4.17	-6.82
		B2	1.17	1.17	0.12	4.40	-6.57
40m Grid 2	40m Grid		0.84	0.71	0.44	3.84	-5.33
	56.56m G	1	1.01	0.91	0.44	5.04	-5.75
		2	0.99	0.88	0.45	4.74	-5.33
	80m Grid	F1	1.29	1.23	0.42	6.81	-6.52
		F2	1.28	1.21	0.42	4.92	-5.87
		B1	1.25	1.16	0.46	6.07	-5.36
		B2	1.21	1.14	0.41	5.34	-5.87

N.B. F=Forward, B=Backward sets of measurements. The same annotation will be used in other tables.

Table 9.7 Accuracy results from gridded data sets for the Sohnstetten and Spitze areas

Area	Data Sets			RMSE	SD	MEAN	+Emax	-Emax
	20 metre Grid			0.57	0.56	-0.11	3.03	-3.33
S	28.28m Grid		1	0.86	0.86	-0.02	4.03	-3.66
O			2	0.88	0.88	-0.02	4.95	-4.06
H	40 m Grid		F1	1.57	1.56	0.19	5.95	-5.42
N			F2	1.54	1.53	0.12	7.61	-9.14
S			B1	1.34	1.34	0.08	5.71	-7.82
T			B2	1.35	1.35	0.10	6.29	-6.37
E	56.56m Grid	F1	1	2.42	2.37	0.46	11.16	-5.69
T			2	2.34	2.33	0.29	10.68	-10.62
T		F2	1	2.59	2.57	0.37	11.16	-11.11
E			2	2.27	2.26	0.27	10.52	- 6.54
N		B1	1	2.34	2.31	0.33	11.97	- 9.09
			2	2.49	2.43	0.54	11.35	- 8.05
		B2	1	2.46	2.42	0.42	12.18	- 6.57
			2	2.54	2.47	0.59	15.08	- 6.06
	10m Grid			0.22	0.21	0.07	1.44	-2.26
S	14.14 m Grid		1	0.29	0.28	0.07	1.70	-3.06
P			2	0.28	0.28	0.06	1.93	-1.77
I	20 m Grid		F1	0.39	0.39	0.06	2.79	-3.49
T			F2	0.34	0.34	0.05	3.17	-1.91
Z			B1	0.37	0.36	0.06	3.05	-3.02
E			B2	0.36	0.36	0.07	2.92	-1.93

Table 9.8 Occurrence frequencies of large residuals from gridded data sets of the Uppland area

Grid	Data Set		> 2 * SD		> 3 * SD		> 4 * SD	
			No.	%	No.	%	No.	%
Grid 1 + 2	28.28m Grid		112	4.8	22	0.95	7	0.3
40m Grid 1	40m Grid 1		119	5.1	21	0.9	7	0.3
	56.56m Grid	1	127	5.5	25	1.1	6	0.3
		2	130	5.6	25	1.1	7	0.3
	80m Grid	F1	130	5.6	31	1.3	8	0.3
		F2	135	5.8	26	1.1	10	0.4
		B1	123	5.3	24	1.1	7	0.3
		B2	138	6.0	27	1.2	3	0.1
40m Grid 2	40m Grid 2		115	5.0	32	1.4	7	0.3
	56.56m Grid	1	110	4.8	37	1.6	10	0.4
		2	111	4.8	30	1.3	10	0.4
	80m Grid	F1	120	5.2	30	1.3	10	0.4
		F2	133	5.7	33	1.4	8	0.3
		B1	128	5.5	28	1.2	11	0.4
		B2	134	5.8	36	1.1	9	0.3

Where, "> N * SD" means the residuals more than or equal to N times the standard deviation distant from the mean; No. denotes the number of large residuals and % the percentage occurrence frequency. The same annotation will be used in the other tables showing occurrence frequencies.

Table 9.9 Occurrence frequencies of large residuals from gridded data sets of Sohnstetten and Spitze areas

Grid	Data Set		> 2 * SD		> 3 * SD		> 4 * SD	
			No.	%	No.	%	No.	%
Sohnstetten	20m Grid		106	5.6	33	1.7	15	0.8
	28.28m Grid	1	119	6.3	30	1.6	8	0.4
		2	106	5.6	35	1.8	14	0.7
	40m Grid	F1	131	6.9	21	1.1	0	0.0
		F2	133	7.0	42	2.2	8	0.4
		B1	115	6.1	24	1.3	8	0.4
		B2	120	6.3	25	1.3	5	0.3
	56.56m Grid	F1 1	117	6.2	16	0.8	2	0.1
		F1 2	115	6.1	28	1.5	7	0.4
		F2 1	111	5.9	31	1.6	5	0.3
		F2 2	115	6.1	27	1.4	5	0.3
		B1 1	121	6.4	41	2.2	9	0.5
		B1 2	113	6.0	22	1.2	3	0.2
		B2 1	121	6.4	30	1.6	8	0.4
		B2 2	106	5.6	33	1.7	9	0.5
Spitze	10m Grid		106	5.0	48	2.3	31	1.5
	14.14m Grid	1	113	5.3	54	2.6	26	1.2
		2	131	6.2	61	2.9	25	1.2
	20m Grid	F1	89	4.2	48	2.3	30	1.4
		F2	124	5.9	61	2.9	28	1.3
		B1	114	5.4	61	2.9	38	1.8
		B2	130	6.1	58	2.7	29	1.4

9.5.3 Accuracy of the DTM derived from composite data sets

As discussed in Chapter 4, there are two types of composite data set, i.e. (i) regularly gridded data with a fixed grid interval plus selected feature-specific points, and (ii) gridded data with varying grid intervals plus selected feature-specific points. The data sets used in this test belong to the former type.

For the data sets used in this test, the form lines and/or break lines have been sampled with a great density. For example, the number of points contained in these lines for the Uppland area is about 1,750 while the total number of points produced by the two 40m grids is only 3,638. To give another example, for the data set for the area of Sohnstetten, the total number of points provided by these break lines and form lines is about 1,550, while that provided by the 20m grid is only 1,716. Therefore, during the process of triangulation, about 150 points from the form and break lines for each area were filtered out automatically by the triangulation program itself because they were considered to be duplicated.

The composite data sets should be providing a higher fidelity in terms of representing the topography of the terrain surface than the regularly gridded data sets. That is to say, the accuracy of the DTM formed from a composite data set should be higher than that resulting from the use of regular gridded data only. The results from each of these composite data sets are given in Tables 9.10 and 9.11. The diagrammatic representation of these test results is shown in Figs.9.12(b), (d) and (f). The occurrence frequencies of the large residuals for each of the test areas are also recorded and given in Tables 9.12 and 9.13.

Table9.10 Accuracy results from the composite data sets for the Uppland area

Grid No.	Data Set		RMSE	SD	MEAN	+Emax	-Emax
Grid 1 + 2	28.28m G + P		0.64	0.59	0.24	2.41	-5.78
40m Grid 1	40m Grid 1 + P		0.67	0.66	0.10	3.33	-6.42
	56.56m G + Points	1	0.71	0.70	0.10	2.69	-6.42
		2	0.71	0.71	0.09	3.29	-6.51
	80m Grid + Points	F1	0.84	0.83	0.10	3.15	-5.91
		F2	0.82	0.81	0.11	4.68	-5.88
		B1	0.79	0.78	0.13	3.19	-6.47
		B2	0.79	0.78	0.12	3.74	-6.43
40m Grid 2	40m Grid 2 + P		0.72	0.63	0.35	2.41	-5.33
	56.56m G + Points	1	0.77	0.70	0.32	3.48	-5.75
		2	0.74	0.67	0.31	2.70	-5.33
	80m Grid + Points	F1	0.86	0.82	0.24	4.11	-6.52
		F2	0.83	0.77	0.31	3.46	-5.42
		B1	0.83	0.77	0.30	3.76	-5.36
		B2	0.83	0.77	0.32	4.19	-5.16

Where, the term points refers to the feature specific points and those located along form lines and break lines. Also G=Grid and P=Points. The same annotations will be used later in other tables.

Table 9.11 Accuracy results from the composite data sets for the Sohnstetten and Spitze areas

Test Area	Data Set		RMSE	SD	MEAN	+Emax	-Emax
Sohnstetten	20m Grid + P		0.43	0.40	-0.15	1.68	-2.55
	28.28m G + Points	1	0.55	0.53	-0.14	2.00	-3.63
		2	0.58	0.56	-0.13	2.40	-3.30
	40m Grid + Points	F1	0.79	0.78	-0.15	2.82	-3.00
		F2	0.78	0.77	-0.15	2.90	-4.34
		B1	0.78	0.77	-0.14	3.23	-3.75
		B2	0.78	0.76	-0.15	2.67	-4.63
	56.56m Grid + P	F1 1	1.08	1.08	-0.12	5.12	-5.19
		F1 2	1.09	1.07	-0.22	3.37	-5.19
		F2 1	1.09	1.07	-0.19	3.79	-6.27
		F2 2	1.09	1.08	-0.18	4.45	-5.51
		B1 1	1.07	1.06	-0.16	4.59	-5.59
		B1 2	1.10	1.08	-0.20	3.73	-4.98
		B2 1	1.09	1.07	-0.23	4.38	-5.22
		B2 2	1.12	1.12	-0.10	5.43	-3.73
Spitze	10m Grid +P		0.16	0.14	0.07	0.87	-0.79
	14.14m G + Points	1	0.17	0.16	0.07	0.88	-2.71
		2	0.17	0.15	0.07	0.88	-2.66
	20m Grid + Points	F1	0.175	0.16	0.05	0.88	-2.38
		F2	0.18	0.17	0.06	0.88	-2.66
		B1	0.174	0.16	0.06	0.88	-2.27
		B2	0.174	0.16	0.06	0.88	-2.72

Table 9.12 Occurrence frequencies of large residuals from the composite data sets for the Uppland area

Grid	Data Set		> 2 * SD		> 3 * SD		> 4 * SD	
			No.	%	No.	%	No.	%
Grid 1 + 2	28.28m Grid		101	4.3	18	0.8	4	0.2
40m Grid 1	40m Grid + P		92	4.0	16	0.7	7	0.3
	56.56m G + Points	1	102	4.4	20	0.9	5	0.2
		2	104	4.5	24	1.0	4	0.2
	80m G + Points	F1	112	4.8	36	1.5	10	0.4
		F2	120	5.2	32	1.4	9	0.4
		B1	116	5.0	30	1.3	5	0.2
		B2	114	4.9	29	1.2	5	0.2
40m Grid 2	40m Grid + P		110	4.8	18	0.8	4	0.2
	56.56m G + Points	1	115	5.0	28	1.2	5	0.2
		2	119	5.1	21	0.9	2	0.1
	80m G + points	F1	111	4.8	24	1.0	9	0.4
		F2	124	5.3	30	1.3	5	0.2
		B1	133	5.7	26	1.1	3	0.1
		B2	115	5.0	31	1.3	12	0.5

Table 9.13 Occurrence frequencies of large residuals from the composite data sets for the Sohnstetten and Spitze areas

Area	Data Set		> 2 * SD		> 3 * SD		> 4 * SD	
			No.	%	No.	%	No.	%
Sohnstetten	20m Grid + P		109	5.8	26	1.4	9	0.5
	28.28m G + Points	1	107	5.7	18	1.0	5	0.3
		2	104	5.5	31	1.6	8	0.4
	40m Grid + Points	F1	122	6.4	24	1.3	0	0.0
		F2	125	6.6	24	1.3	3	0.2
		B1	110	5.8	30	1.6	6	0.3
		B2	105	5.5	28	1.5	6	0.3
	56.56 m Grid + P	F1 1	130	6.3	24	1.3	7	0.3
		F1 2	112	6.3	24	1.3	5	0.3
		F2 1	128	6.3	22	1.3	4	0.3
		F2 2	97	6.3	26	1.3	6	0.3
		B1 1	116	6.3	36	1.3	13	0.3
		B1 2	125	6.3	18	1.3	3	0.3
		B2 1	129	6.3	30	1.3	4	0.3
		B2 2	107	6.3	30	1.3	9	0.3
Spitze	10m Grid + P		101	4.7	39	1.8	16	0.8
	14.14m G + Points	1	84	4.0	30	1.4	8	0.4
		2	87	4.1	33	1.6	15	0.7
	20m G + Points	F1	96	4.5	33	1.6	9	0.4
		F2	97	4.6	33	1.6	7	0.3
		B1	88	4.2	39	1.8	16	0.8
		B2	96	4.5	31	1.5	7	0.3

9.6 Analysis of test results

The test results have been given in the previous section and some analysis will be given in this section. This will include a descriptive analysis of the accuracy results obtained from both the regularly gridded data sets and the composite data sets; a regression analysis of the obtained accuracy figures; and a descriptive analysis of the occurrence frequencies of large residuals for these different data sets.

9.6.1 Descriptive analysis of the accuracy results

(1). Analysis of the accuracy results for the Uppland area

From Table 9.6 giving the accuracy results obtained from the **regular gridded data sets** for the Uppland area, it can be seen that the mean values and the RMSE values of the DTM residuals obtained from the first 40m grid for the Uppland area are quite different to those from the second grid, but the SD values are quite similar. This might suggest that there is a systematic shift between these two data sets. The value of this shift is about 0.30m. Evidence showing such a trend is also given by the mean resulting from the grid data set with the 28.28m interval which was generated by adding the two 40m grids together. This gives a value of 0.26m which is almost equal to the average of the two means (0.10m and 0.44m) resulting from the two 40m grids.

For the **composite data set** for the Uppland area, the accuracy results are given in Table 9.10. From this table, again, a constant shift between the two 40m grids can be observed. However, the value of the shift becomes smaller - about 0.2m - in this case, when the feature-specific points are added to the gridded data. The amount of reduction is about 0.1m. Another interesting point arising from this change is that the value of the mean resulting from the second 40m grid is reduced by 0.12m. By contrast, the mean resulting from the first 40m grid is increased very little by 0.03m (which is insignificant).

Also it can be found from inspection of Figs.9.12(a) and (b) that the RMSE and SD values increase with an increase in the sampling interval. In both cases, the trend looks very linear. The only difference between these two trends is that the speed of decrease in accuracy is faster in the case of regular gridded data than when composite data is used.

It is also of interest to notice that, in this test, the accuracy figures from the data sets measured in the forward direction are almost the same as those of the corresponding data sets measured in the backward direction.

(2). Analysis of the accuracy results for the Sohnstetten and Spitze areas

The accuracy results from the gridded data for the Sohnstetten area are listed in Tables 9.7 and 9.11. From Table 9.7, it can be found that, in general, the mean error values increase with the increase in the grid interval. The magnitudes of the extreme errors (i.e. the positive and negative maximum errors) also show a similar tend.

The reason for these phenomena may be found from the test area itself. It is not difficult to notice that the area covered by the check grid lies, in the most part, along and on either side of a ravine. It is not difficult to imagine that, with the increase in the grid interval, the DTM surface which is constructed linearly from the gridded data has been lifted up over the ravine area. If this is the case, then the magnitude of the errors (positive in this case) in the points lying on the DTM surface will increase. The consequence is an increase in the magnitude of the positive maximum errors. Also in this area, a small part covers a ridge line. As in the case of the ravine area, the DTM surface covering this small area will be lowered when a larger grid interval is used. This would result in an increase in the magnitude of the negative errors, and thus of the maximum value. However, since most of the area lies on either side of the ravine line, so the increase in the sum of the positive errors will be far larger than that of the negative errors. Thus the resulting mean value increases in a positive sense with the increase in the grid interval.

Fig.9.13 is the plot of the DTM errors for the 20m grid data for the Sohnstetten test area. The size of a circle indicates the magnitude of the DTM error. The red circles indicate the positive errors and the blue circles denote the negative errors. From this diagram, it can be seen clearly that the positive errors are almost all located along the bottom of the ravine while the largest negative errors are located along the ridge lines (e.g. see top left on the diagram). This diagram provides a very strong back up to the analysis given above. Another strong evidence to back up the reasoning given above is provided by the fact (see Table 9.11) that, after adding the points measured along the ravine lines and ridge lines, the magnitude of the extreme errors is significantly reduced and the mean value is kept almost constant.

It is of interest to note that the two extreme values of the DTM errors for the grid data for the Sohnstetten test area appear very large when the grid interval reaches a value of 56.56m. The values for the maximum errors range from 10.52m to 15.08m. These values are larger than 0.67 per mil of the flying height and about 10% of the height range. At the first look, they seem enormous. However, they can by no means be considered as gross errors and they are due to the inherent nature of

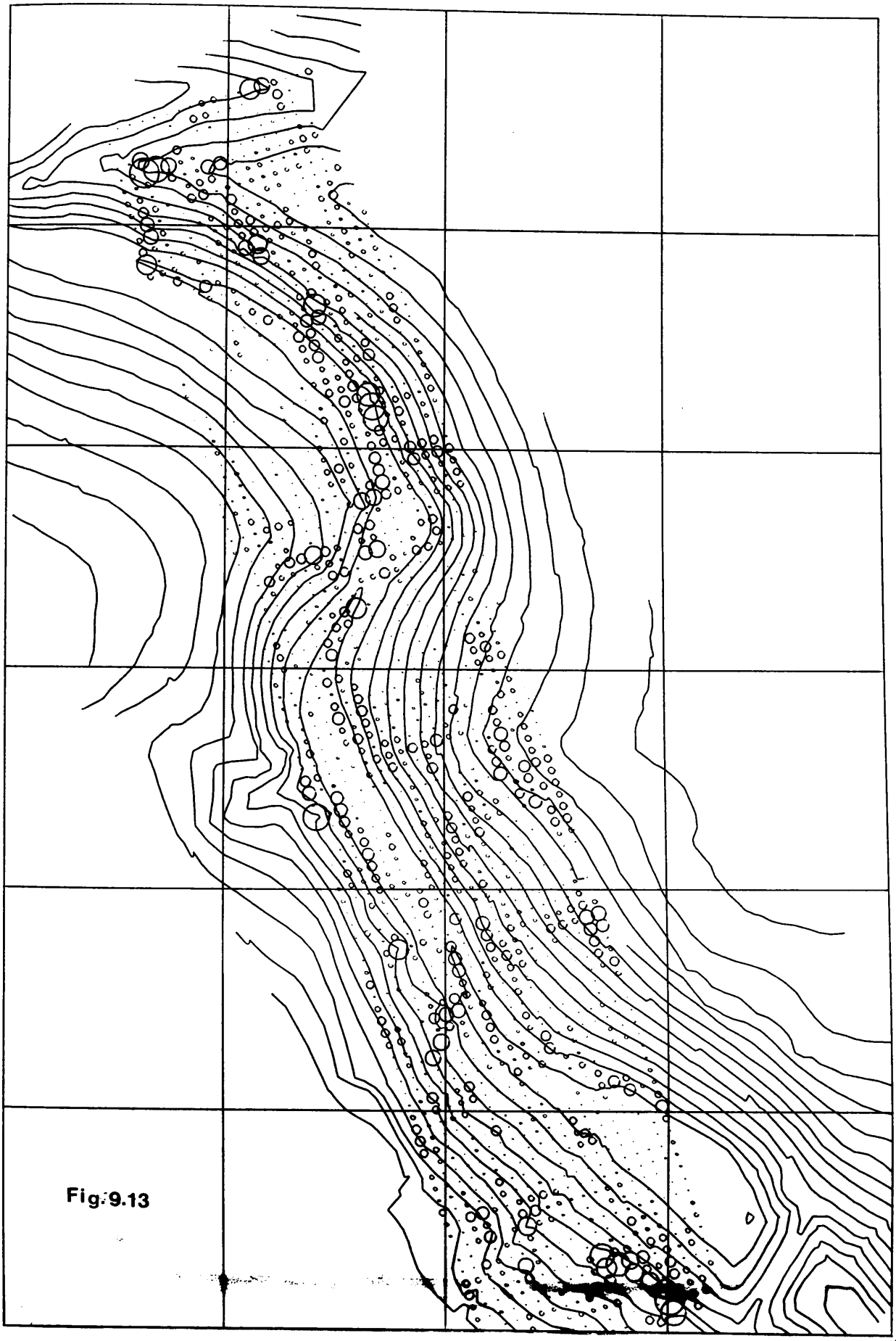


Fig.9.13

Fig.9.13 Distribution of DTM errors for the 50m grid-data for the Sohnstetten area

(red circles: positive errors; blue circles: negative errors)

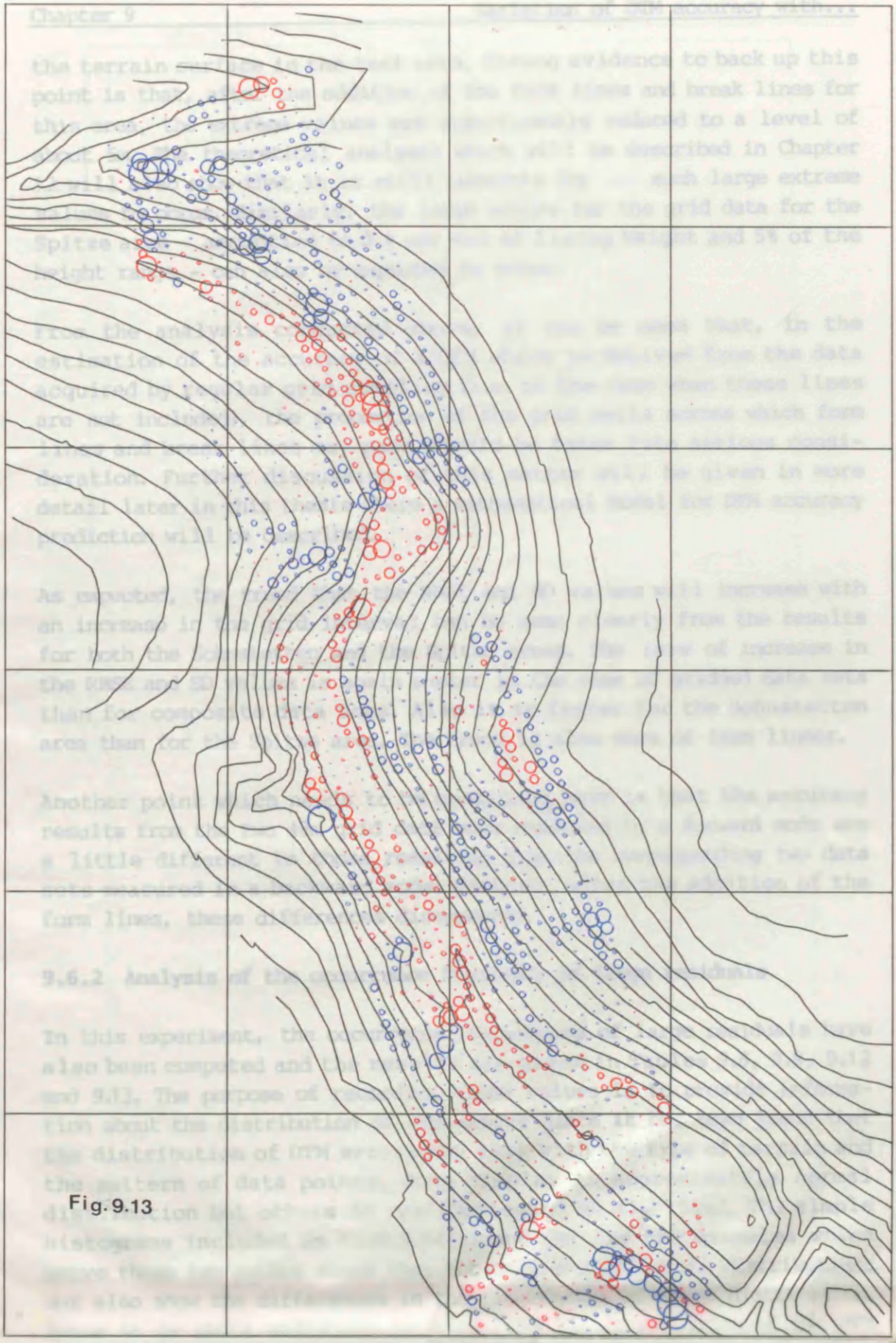


Fig.9.13

Fig.9.13 Distribution of DTM errors for the 20m gridded data for the Sohnstetten area
(red circles: positive errors; blue circles: negative errors)

the terrain surface in the test area. Strong evidence to back up this point is that, after the addition of the form lines and break lines for this area, the extreme values are significantly reduced to a level of about 5m. The theoretical analysis which will be described in Chapter 12 will also show that it is still possible for such large extreme values to occur. Similarly, the large errors for the grid data for the Spitze area - amounting to 0.4 per mil of flying height and 5% of the height range - can also be expected to occur.

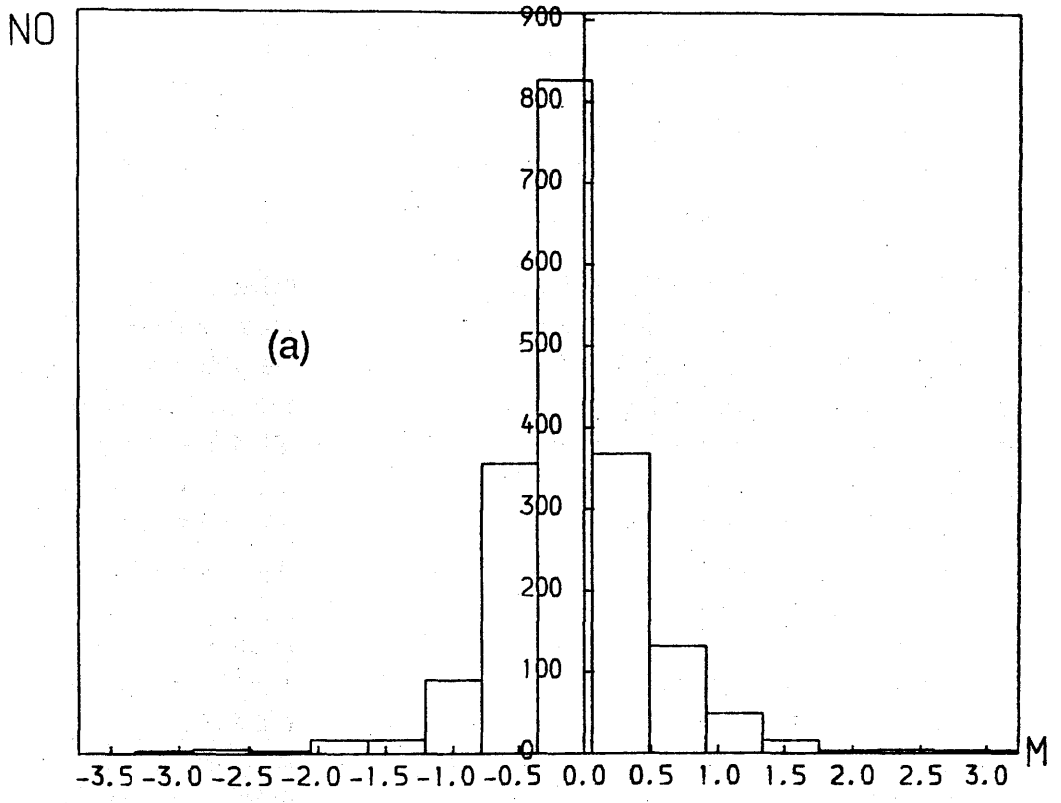
From the analysis conducted above, it can be seen that, in the estimation of the accuracy of a DTM which is derived from the data acquired by regular grid sampling (i.e. in the case when these lines are not included), the proportion of the grid cells across which form lines and break lines may pass should be taken into serious consideration. Further discussion of this matter will be given in more detail later in this thesis where a mathematical model for DTM accuracy prediction will be described.

As expected, the trend that the RMSE and SD values will increase with an increase in the grid interval can be seen clearly from the results for both the Sohnstetten and the Spitze areas. The rate of increase in the RMSE and SD values is again greater in the case of gridded data sets than for composite data sets. Also it is faster for the Sohnstetten area than for the Spitze area. The trend is also more or less linear.

Another point which needs to be mentioned here is that the accuracy results from the two 40m grid data sets measured in a forward mode are a little different to those resulting from the corresponding two data sets measured in a backward mode. However, after the addition of the form lines, these differences disappeared.

9.6.2 Analysis of the occurrence frequency of large residuals

In this experiment, the occurrence frequencies of large residuals have also been computed and the results are given in Tables 9.8, 9.9, 9.12 and 9.13. The purpose of recording these values is to provide information about the distribution of DTM errors since it has been found that the distribution of DTM errors may vary with the type of terrain and the pattern of data points. Some display an approximately normal distribution but others do not (Torlegard et al, 1986). The simple histograms included in Figs.9.14(a) and (b) are two examples which prove these two points since they not only show different distributions but also show the differences in the frequencies of error distribution. Since it is still difficult to visualize the distribution of the DTM errors, it is, therefore, necessary to undertake a detailed examination of the occurrence frequencies of large residuals.



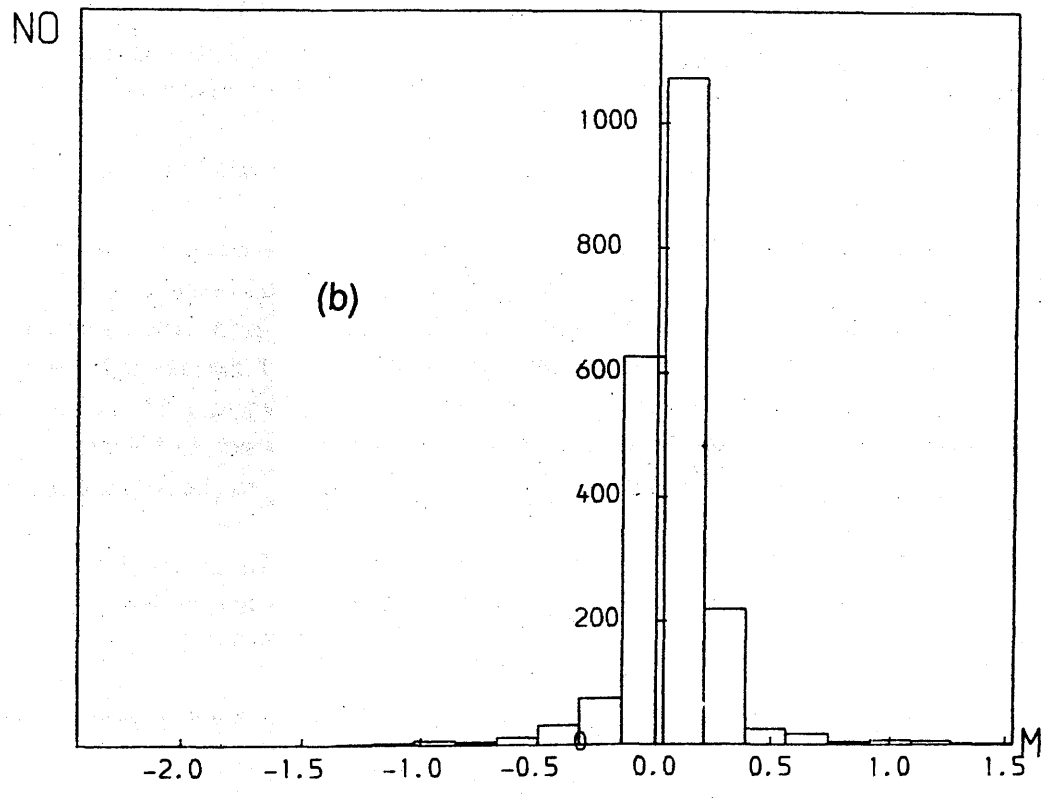


Fig.9.14 Histogram for the distribution of the DTM errors
(a). Results from composite data (20m grid) for the Sohnst area
(b). Results from regular gridded data (10m grid)
for the Spitze area

The occurrence frequencies of large residuals may also vary with terrain type and with the pattern of the data points. One would expect that the occurrence frequencies of large residuals will be much higher for broken terrain than for smooth terrain, if the data for the break lines and form lines are not used for surface reconstruction.

These large residuals have been classified into three classes, i.e. those larger or equal to 2 times SD (standard deviation), 3 times SD and 4 times SD. From the test results obtained from this study, it was found that the frequency of the residuals larger than 2 times SD is smaller than 7.0%. For those residuals larger than 3 times SD, it is 2.0%. Finally, the frequency figure is much smaller than 1.0% for those errors larger than 4 times SD in all cases, except for the gridded data set of Spitze area where terrain discontinuities exist.

Therefore, from this series of tests, it can be stated that the vast majority of DTM errors are smaller than 4 times SD (the standard deviation). Comparing this with the normal distribution, it can also be found that the probability with which DTM errors fall within the range from -4xSD to 4xSD from the mean is very approximately equal to the probability with which random errors of normal distribution fall within the range of 3 times the corresponding SD. This is a very important conclusion for the development of mathematical models for DTM accuracy prediction which will be discussed later in this thesis.

9.6.3 Regression analysis of the accuracy results

As stated previously, the main purpose of this experimental test is to obtain information about the general relationship between sampling interval and the accuracy of the resulting digital terrain model. A discussion about the general impression given by the results has been given in the previous section (9.6.1). In this section, a regression analysis will be carried out to obtain some quantitative results. The procedure used in this study is as follows:

i). First of all, a mathematical model is selected based on experience. Thus in practical terms, the selected model is usually an empirical model.

ii). Then the experimental test results are used to compute the parameters (coefficients) of this model.

iii). Finally, the adequacy of this model is examined by the value of its correlation coefficient and/or through variance analysis.

The mathematical model used in this analysis is as follows:

$$VAR(DTM) = a^2.VAR(PMD) + b.(Dx.tanA)^2 \qquad (9.1)$$

Where, a and b are two coefficients; VAR(PMD) denotes the variance of PMD (the photogrammetrically measured data); A refers to the typical slope angle in the area; Dx is the sampling interval (the grid interval); and VAR(DTM) denotes the variance of the resulting DTM. This is, in fact, a model which is similar to that suggested by Ackermann (1979).

It can be seen that, for the same data set, the first term of the right side of Equ.9.1 is a constant; and $b.tan^2A$ will also be a constant for the same area. Therefore, in the case of this study, this model can simplified as follows:

$$VAR(DTM) = c + d.Dx^2 \qquad (9.2)$$

where c and d are the coefficients. The regression results are shown in Table 9.14.

For the Uppland area, due to the constant shift between two 40m grids, the SD values were used instead of the RMSE values. For the accuracy results for the Spitze and Sohnstetten areas, the mean values of the residuals are comparatively small, therefore, the SD values are almost equal to the RMSE values. Thus, either of these two values will do. Actually, the RMSE values have been used in this analysis.

Table 9.14 Regression results for the coefficients "c" and "d"

Test Area	Data Set	c	d	r
A. Uppland	Gridded	0.2575	0.0001775	0.983
	Composite	0.3335	0.0000459	0.953
E. Sohnstetten	Gridded	-0.8269	0.0019505	0.990
	Composite	0.0253	0.0003719	0.999
F. Spitze	Gridded	0.0240	0.0002749	0.955
	Composite	0.0251	0.0000151	0.890

where, r is the correlation coefficient.

By substituting the value of VAR(PMD) into Equ.(9.1), then the value of the coefficient "a" can be obtained for each of the different test areas. The values of "a" should lie in between the values 0.0 and 1.0. The reason is very obvious, in that, if any parameter in the second term on the right side of Formula (9.1) is zero, then this term will be

zero. In which case, the errors in the DTM points are simply propagated directly from the source data points. In this case, the DTM height is the mean (maybe weighted) of the reference points which have been used for interpolation. The accuracy of the mean value is, of course, higher than the accuracy of the reference points. Thus "a" should lie within the range 0.0 to 1.0.

Table 9.15 Regression results for the coefficient "a"

	Uppland	Sohnstetten	Spitze
Gridded	0.7574	-----	1.9365
Composite	0.8620	0.9990	2.8228

It was found that this model is more or less suitable for the results from the data sets of the **Uppland area** and the composite data set of the Sohnstetten area which are set out in Table 9.15. However, the results for the gridded data set of the **Sohnstetten area** show that the value of the variance of the DTM errors from that data set is proportional to the sampling interval (grid cell in this case) to a power greater than 2. Of course, the results from the regression process are very sensitive to even a small variation. However, in any case, the results obtained from the gridded data set for the Sohnstetten area strongly indicate that quite different mathematical models should be used for gridded data sets and composite data sets.

It would seem that the results from the data sets for the **Spitze area** indicate that this model does not fit the experimental data. One of the reasons could be that too few sampling intervals were used for the analysis, therefore the results which were obtained are very unreliable.

Of course, if one likes, one can try different mathematical models to see how will they fit these test results. However, this is not the primary objective of this project. Realistically, even if a mathematical model does fit the practical data very well, it does not necessarily mean that this model expresses the fundamental or inherent relationship between the variables under investigation. On the other hand, a model which does not fit a specific set of experimental data is not necessarily inadequate. Thus, as has been discussed in Chapter 2, for a particular mathematical model, the important thing which one should do is to estimate how wrong it is instead of just rejecting it. Therefore, for the time being, the only conclusion which can be reached is that the model expressed by Formula (9.1) does apply in some cases but not always. Further discussion about this matter will be given in Chapter 12 where a deeper investigation into the mathematical models

which can be used for DTM accuracy will be described.

9.7 Conclusions

From the experimental tests described in this chapter, the following concluding remarks can be made:

i). The accuracy of a DTM formed from photogrammetrically measured data is highly correlated with the **sampling interval** (grid interval) if only gridded data is used.

ii). When **feature-specific points** are added to the data set, the accuracy of the DTM can be improved. This improvement is greater if the sampling interval is large. With a small sampling interval, this effect may not be significant.

iii). Large residual errors do occur but the **occurrence frequencies** of these residual errors being greater than 4xSD is usually very small. With the inclusion of feature-specific points (including points measured along break lines and form lines), the magnitude of large residuals can be reduced. If the area has many break lines and terrain discontinuities, this reduction could be significant.

iv). It was also found that the accuracy of DTMs is correlated with the **slope angle** of the terrain surface. In those areas with steeper slopes, the RMSE and SD values increase with a faster speed.

v). The accuracy results obtained from **two data sets of the same area** may be quite different, even if the sampling interval is the same for both of them. Therefore, it is impossible to be definite about the accuracy value which can be obtained from a data set employing a given sampling interval. However, an approximate value or a range of values can be given for the accuracy.

In this chapter, an experimental test deliberately designed to establish the variation of DTM accuracy with sampling interval has been presented. The results of this test will be used to validate the mathematical models developed in this project which will be presented in Chapter 12. As has been described previously, the scale of the aerial photography used in this test ranges only from 1:4,000 to 1:30,000. In the next chapter, an experimental test on the accuracy of the DTMs derived from the data measured from the space photography taken from the Metric Camera will be presented as an extension of this study.

Chapter 10

Accuracy of the DTMs Derived from Metric Camera Photography

Chapter Ten

Accuracy of the DTMs Derived from Metric Camera Photography

10.1 Introduction

In Chapter 8, the first of a series of tests on DTM data acquired through photogrammetric measurement was reported upon. This was concerned with data acquired using digital correlation (i.e. image matching) techniques and in accordance with the characteristics of this data, the tests and discussion concentrated on the matter of data filtering and gross error detection with the objective of improving the accuracy of the measured data. In Chapter 9, an experimental investigation into the accuracy of the DTMs derived from source data acquired from aerial photography using standard manual photogrammetric methods of measurement based on the use of an analytical plotter has been presented.

In this chapter, a further test of the accuracy of a DTM, in this case derived from space photography, will be described. In this context, as has been pointed out in Chapter 4, only a few types of space photography are in experimental use. In this particular project, the accuracy of a DTM derived from the data sets acquired from one of these systems - the Metric Camera (MC) - has been investigated to extend the range of the photographic scales and flying heights over which DTM data accuracy can be obtained and evaluated. In addition, a different method of measurement - dynamic profiling - has been employed, again with a view to extending the range of data acquisition methods being evaluated.

10.2 Background to this study

10.2.1 Background to Metric Camera Experiment

The main objective of the Metric Camera (MC) experiment was to obtain high resolution space photography for topographic and thematic mapping and for map revision at scales of 1:50,000, 1:100,000, 1:250,000 and smaller, since it was believed that there is a great demand for such maps in both developed and developing countries. In developed countries, there is a need for continuous updating (revision) of such maps over a cycle of 2 to 3 years, which is often difficult to implement by conventional means. In some developing countries, especially those in South America and Africa, many countries do not have anything like complete coverage even at these small scales. As a result, the two main space cameras - the Metric Camera (MC) and the NASA Large Format Camera (LFC) - have been used extensively to carry out experimental tests using a wide variety of techniques to produce

maps by both conventional analogue and digital photogrammetric instrumentation and techniques.

Since the distribution of MC photographs in January 1984, a lot of work has already been carried out for many different purposes such as topographic mapping, orthophoto production, etc. However, it is important to test the accuracy of digital terrain models derived from MC photography in order that the effectiveness, flexibility and economy of applying the MC photography to produce DTMs and other related products can be fully evaluated.

10.2.2 Previous work

Some work in this area has already been done by one or two investigators, e.g. that carried out by Ackermann and Stark (1985), who have presented their test results at the Metric Camera Workshop held at Oberpfaffenhofen in 1985. Two different pairs of MC photographs covering an area of fluvio-glacial morphology with height variations from 360m to 600m, located in South Bavaria in West Germany, were used and a DTM produced from each pair. The forward overlap was 80%. In fact, only a small area of 5km x 25km was measured with a Zeiss Oberkochen Planicomp C-100 analytical plotter. Each model was absolutely orientated using 10 control points. These control points were well-defined terrain details such as the crossings or junctions of highways or rivers, etc. The planimetric and height coordinates of these points were interpolated graphically from the 1:25,000 scale West German topographic maps covering the test area. The influence of Earth curvature was automatically corrected for during the measurement of the stereo-models with the analytical plotter. The mean y-parallax after relative orientation was about 8um and the mean of the residual errors after absolute orientation was 11m in the X and Y, and 15m in the Z direction.

The elevation data were acquired in the form of a regular grid with a sampling interval of 100m. A total of 12,500 (50x250) grid points were measured, and a futher 160,000 DTM points were interpolated from these 12,500 measured points using the linear prediction method implemented in the SCOP package. By comparing the two sets of DTM data, a relative accuracy of 22m was obtained. A further 100 spot heights taken from the 1:25,000 scale topographic maps were used as check points and the result of comparing the DTM elevation values with those of the check points gave a root mean square error (RMSE) of about $\pm$24m. The maximum value of the residuals was 75m.

Elmhorst and Muller (1988) also carried out an experimental test on the accuracy of DTMs derived from a pair of Metric Camera photographs with B:H = 0.3. Their test area was a hilly area located near the Bavarian

Alps in West Germany. The size of the test area was about 6.3km x 5.6km. The data was measured using a Zeiss Oberkochen Planicomp C-100 analytical plotter. The composite sampling method was used. The interval for the grid was 100m. In total, 3,648 (64x57) grid nodes were measured. In addition, a total of 208 surface-specific points and points lying on surface-specific lines were also measured. The TASH program which was developed in the University of Hannover was used for the interpolation of the DTM data. The positions and elevations of 92 spot points were digitised from the 1:10,000 and 1:25,000 scale topographic maps and used as check points, against which the elevation values derived from the DTM points were checked. An RMSE value of $\pm$42.3m was obtained. The corresponding maximum error was 103.5m.

These results are substantially much poorer than those of Ackermann and Stark, whose RMSE value for check points represents 24/250 = 0.096 per mil of H. By contrast, Elmhorst and Muller's RMSE value of $\pm$42.3m represents 42.3/250 = 0.17 per mil of H. Nevertheless both sets of RMSE values lie well within the range of 0.1 to 0.2 per mil of H which certainly represents a most acceptable range of RMSE values for heights interpolated from a photogrammetrically measured set of elevation values.

Obviously the results of the tests reported on above are interesting but they also have a somewhat limited applicability because the investigators have selected and tested only a small area - 5km x 25km in the case of Ackermann and Stark and 6.3km x 5.6km in the case of Elmhorst and Muller. In order to obtain more reliable information for practical production, a complete Metric Camera stereo-model was profiled and a very large area has been tested in this present study.

10.3 The Sudan test area

The model used for this new test covers an area of 100km x 180km in the Red Sea Hills area of Sudan. This is a predominantly hilly and mountainous area with a height variation from 300m to 3,000m above sea level. This area can be considered as one of the most undeveloped areas in Sudan so that the cultural features are at the lowest level of density. Therefore only topographic features such as the tops of hills, the junctions of river beds, etc. could be used as control points for absolute orientation.

The whole model covers a complete sheet together with parts of six other sheets of the 1:100,000 scale topographic map of the Sudan. In this study, only the data covering 3 of these sheets (136, 163 and 164) were used for testing due to the fact that no suitable check points were available for the rest of the area.

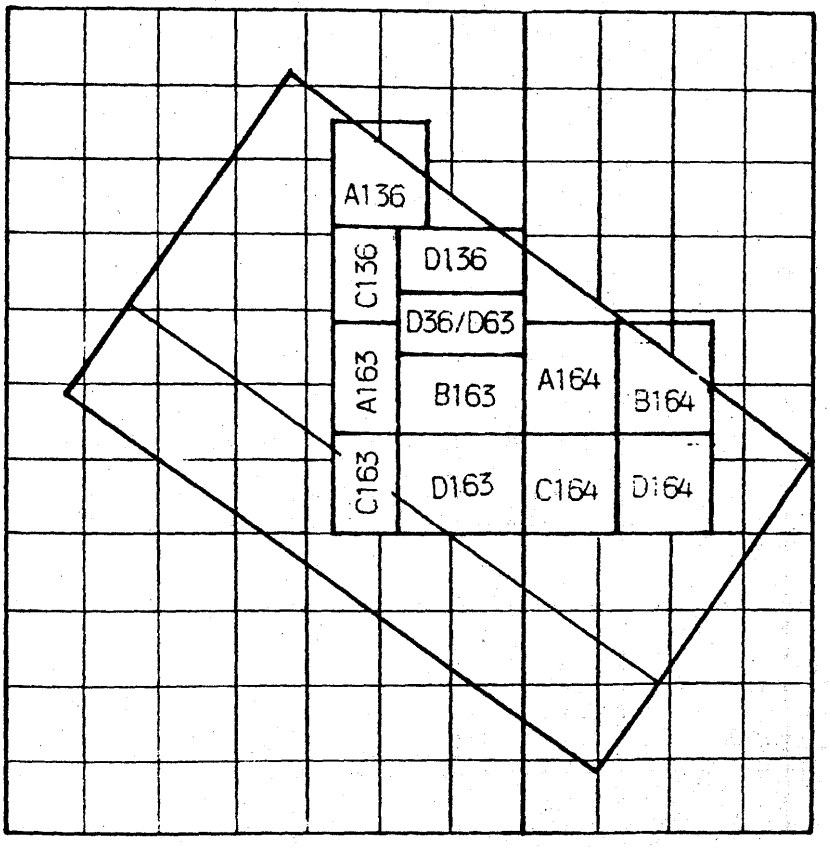
The area used for this testing has also been divided into several sub-areas (Fig.10.1) according to relief range for the purpose of the test since the height varies greatly from place to place. The main morphological parameters for each of these sub-areas are given in Table 10.1.

Table 10.1 Relief Parameters for the Red Sea Hills Test Areas

Area	Contour Range	Slope Range (°)	Typical Slope(°)
A136	540m -- 960m	1 -- 39	< 5, 10
C136	530m -- 1,200m	1 -- 39	< 5, 15
D136	620m -- 1,240m	2 -- 40	about 16
D36/B63	550m -- 720m	1 -- 5	< 5
A163	490m -- 600m	1 -- 15	< 5
B163	550m -- 920m	2 -- 30	about 10
C163	490m -- 920m	1 -- 30	about 12
D163	530m -- 1,120m	5 -- 35	about 20
A164	660m -- 1,040m	1 -- 30	about 12
B164	720m -- 1,560m	5 -- 40	20 -24
C164	650m -- 1,040m	1 -- 35	about 15
D164	780m -- 1,360m	5 -- 40	about 24

Area A136 is, in general, a very flat area. However, in the middle of this area is a small mountainous and hilly area. The slope varies from very small to about 40 degrees. Most of the area of C136 is also very flat, the typical slope being less than 5 degrees. The left half of area D136 is mountainous with a typical slope of about 16 degrees; whereas the right half is a less hilly area in which the height variation is much smaller. Area D36/B63 actually comprises two adjacent areas occuring on Sheets 136 and 163. This is a very flat area with a few hills. Also area A163 is an extremely flat area, with only a few hills. However, B163, C163 and D163 are mountainous and hilly areas with much larger height variations. The area covered by Sheet 164 is, in general, very mountainous, the height variation in this area ranging from about 650m to 1,600m. The largest slope is about 40 degrees. This area is also divided into 4 sub-areas. Among them, A164 is the flattest area and D164 is the steepest one.

Fig.10.1 Sub-areas for the Red Sea Hills test area



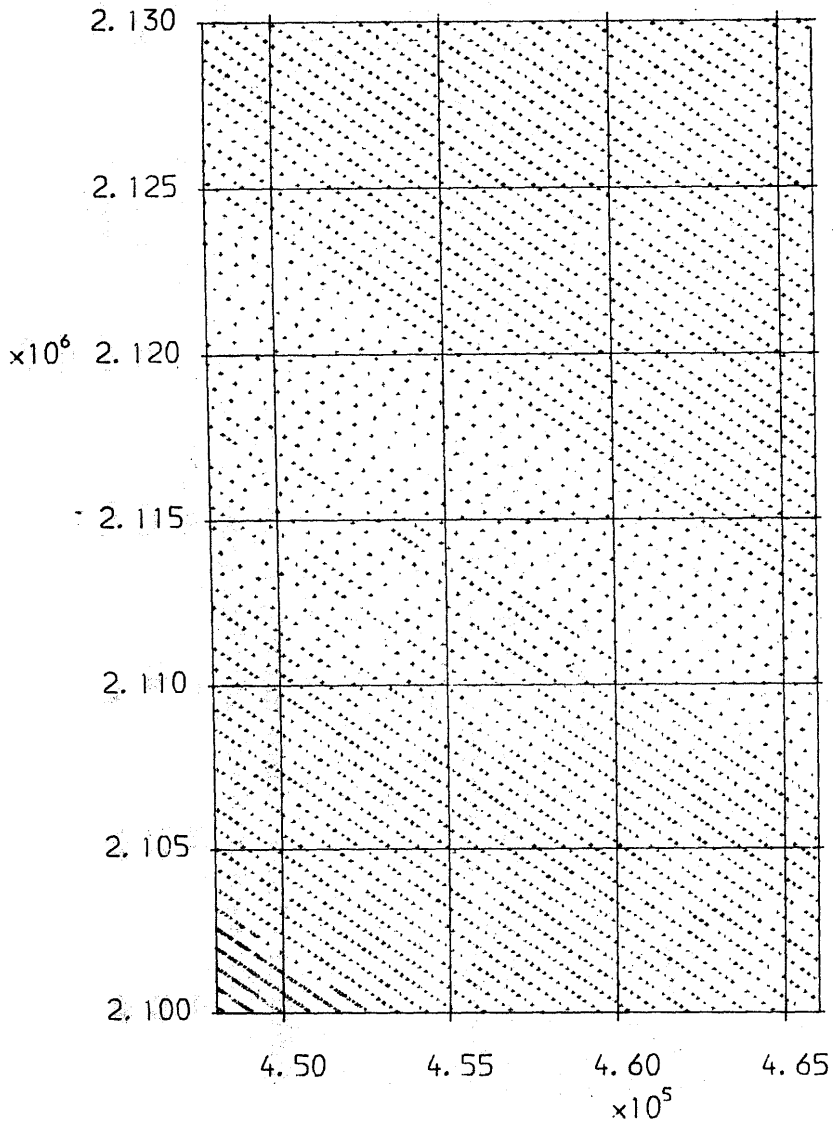


Fig.10.2 Source data for Area U16

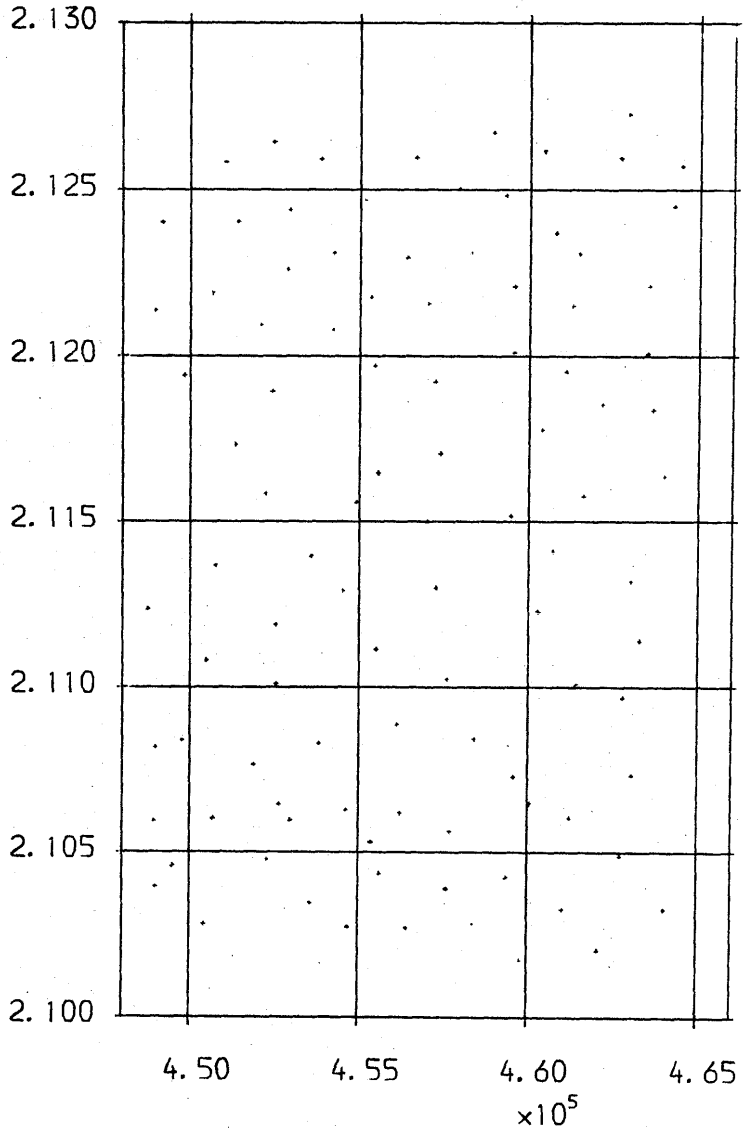


Fig.10.3 Check points for Area U16

10.3.1 DTM source data

The source data was measured directly in the stereo-model by a professional photogrammetric operator and was made available to the present author by courtesy of his reasearch colleague - Dr. A. El-Niweiri. The stereo-model was formed by two MC photographs - Nos.110 and 111. This stereo-pair has also been used for other test work carried out in the Department, e.g. into the planimetric accuracy of topographic features, the accuracy of orthophotographs, etc. (see El-Niweiri, 1988).

The source data was measured on a Kern DSR 11 analytical plotter at the University of York. The whole area was measured at two different times. Profiling in a **dynamic mode** was used instead of a regular grid being measured in a static mode, which is perhaps the more popular method. A method of so-called "distance-controlled profiling" was used to record the data points. A distance interval of 400m was used. This value is the maximum distance in space (essentially a vector) for the measuring marks to move from the previously measured point to the next. In practice, therefore, in hilly areas, the recorded points are much closer in the XY plane than this value of 400m, although in flat areas, the interval between two recorded points will be close to this value. On this model, 182,000 points were measured and recorded in 52 profiles. This took a solid 40 hours of measurement to complete. The distribution of data points is as shown in Fig.10.2.

The detailed information about the orientation of the stereo-model has been described by El-Niweiri (1988) in his Ph.D. thesis. For convenience, a brief quotation is given here. "The inner orientation was carried out using an affine transformation. The largest residual at the fiducial marks in both the x and y directions was 4um. After relative orientation, the standard deviation of all the measured parallaxes was 10um. Absolute orientation was carried out with 14 control points, which are well distributed over the model area. The RMSE values in easting, northing and height which have been obtained at the control points were $\pm$18m, $\pm$18m and $\pm$30m respectively."

10.3.2 Check points

For check points, much larger scale photogrammetric data is highly desirable in order to obtain a very reliable result. However, it is very difficult to do this because the area covered by a single MC stereo-pair is so large. Therefore, as in the case of the two previous West German tests mentioned above, it was decided to use contour data from existing maps instead. That is, a number of points located along the contour lines were digitised from the existing photogrammetrically produced topographic maps for the area and these were then used as

check points for the accuracy testing of the DTM.

The contour maps available in the Department were the 1:100,000 scale topographic maps of the area with a contour interval of 40m in the mountainous areas, 20m in flatter areas, and with 10m complementary contours (form lines) for very flat areas. These maps were compiled from 1:40,000 scale wide angle and 1:60,000 scale super-wide angle aerial photography, both taken from a flying height (H) of 5,500m. Various types of analogue photogrammetric plotter - Wild A7, B8 and Kern PG-2 - were used for the map compilation which was carried out by the UK Directorate of Overseas Surveys as part of a British aid programme to the Sudan. The details have been discussed by Dr. El-Niweiri (1988) in his Ph.D. thesis.

According to the discussion given by Dr. El-Niweiri (1988), averagely, the accuracy of contouring is about $\pm$4.3m. These values of accuracy for the check points were considered as being quite acceptable for use in a test involving MC photography, since the standard deviation of the measuring errors even for stationary (static) measurement could exceed $\pm$20m (Togliatti and Moriondo, 1985; Elmhorst and Muller, 1988) and thus the accuracy of the profile data measured in the dynamic mode employed in this case might well exceed $\pm$20m.

The check points used in this study were, therefore, digitised from the Sudanese 1:100,000 scale topographic maps. During the process of digitisation, an attempt was made to ensure a relatively even distribution of the check points. Fig.10.3 shows an example of the resulting distribution for the C163 area. The numbers of check points used in the tests for different areas are given in Table 10.2.

10.4 Test results and the analysis

The source data was processed using the PANACEA triangulation-based program. Therefore, the points in the profile data were first of all triangulated, then the interpolation of the elevations for the DTM points took place on the surface constructed from the TIN. Linear interpolation within the triangular facets was used for purpose. The results are shown in Table 10.2, where, as usual, the Emax (maximum positive error), -Emax (maximum negative error), Mean (average value of the residuals), SD (the standard deviation of the residuals from the mean) and RMSE (root mean square errors) are all recorded. The symbol "$\pm$" before SD and RMSE values has been omitted simply for convenience. Actually, in some areas, a couple of residuals larger than 100m occured. However, they were considered as gross errors and deleted.

It can be found that all the means are negative, varying from about -7m

to -42m with an overall average of -21.0m. This strongly indicates that there could be either a single overall systematic error or the sum of a few systematic errors affecting the results. Actually, such a systematic error has also been reported by several other investigators carrying out accuracy tests on space imagery. For example, Ackermann and Stark (1985) have reported that in their test, a systematic shift of 15m was found in the measured elevation values. Also a similar experience has been reported recently by investigators at University College London (D. Tait, personal communication). This error(s) may be due to many different factors, e.g the operator keeping the floating marks too far down during operation such as profiling.

Table 10.2 Test results for DTM of the Red Sea Hills area derived from MC photographs

Area	No.of C.P	E_{max}	$-E_{max}$	RMSE	Mean	SD
A136	82	19.5	-51.5	16.1	-8.8	13.5
C136	94	10.1	-68.1	27.2	-23.0	14.5
D136	129	68.5	-80.9	25.1	-6.4	24.3
D36/B63	117	24.9	-60.1	25.9	-19.5	17.0
A163	113	31.3	-47.2	18.6	-6.6	17.4
B163	126	69.1	-90.5	29.9	-18.3	23.7
C163	100	18.8	-72.8	44.6	-41.7	16.6
D163	165	42.5	-93.3	42.0	-33.5	25.3
A164	113	33.2	-72.1	35.4	-30.8	17.5
B164	105	60.7	-80.3	38.2	-30.3	23.4
C164	120	86.5	-75.6	38.0	-32.1	20.3
D164	125	92.7	-99.5	52.2	-26.6	37.2
Overall	1389	92.7	-99.5	34.1	-21.0	26.8

N.B. "No.of C.P" means the number of check points used for this test. The units used for the accuracy figures in this table are metres.

The results given in Table 10.2 were computed after the deletion of these gross errors. The RMSE values vary from $\pm$16.1m to $\pm$52.2m and the SD values from $\pm$13.5m to $\pm$37.2m. The overall RMSE value is $\pm$34.1m and the corresponding SD value is $\pm$26.8m.

In certain mainly flat areas such as A136 and A163, the resulting RMSE values are even smaller than ±20m. These results are very good indeed, especially when one considers the fact that the source data points were measured in a stereo-model formed from a pair of space photographs taken from a flying height (H) of over 250,000m using the dynamic profiling method. Also for two or three other test areas, the resulting RMSE values are about ±25m. Again, these are still very good results, which would be acceptable even with statically measured spot height data.

In general, the results obtained from this testing are very reasonable. The overall RMSE value is about ±0.14 per mil of H. This is well within the normal range of the accuracy expected from photogrammetrically measured data. The largest RMSE value is about ±0.21 per mil of H. This value is still quite acceptable. Indeed, all of these figures fall well within the figure of 0.3 per mil of H suggested by various experimenters with dynamically measured height data as discussed previously in Section 5.6.

The poorest results were obtained from area D164. In this area, unlike the other test areas where either none or only a couple of large residuals exceeded 100m, many large residuals occured exceeding 100m. This may be due to the reason that this is an area over which cloud and haze were present and therefore the measuring conditions were at the poorest level. Another reason why the RMSE value for this area is the largest could well be that this is also the most mountainous area with the steepest slopes.

This test results show clearly that the standard deviation increases with the increase of slope with a very high correlation. On the other hand, the root mean square error (RMSE) does not provide so clear a conclusion because of the existence of the systematic errors which could have resulted from a combination of many different factors. If one was to take 10m as the value for a constant shift for these results, then a RMSE from ±20m to ±35m may be obtained. This test result is quite in accordance with the results which have been obtained by other investigators; at the lower end, the figures can be regarded as very good indeed.

10.5 Concluding remarks

From this intensive test, the following conclusions can be made:

i). In general, the accuracy of the DTM data decreases with the increase of **slope** for the same data sets.

ii). Some kind of systematic errors are very likely to happen. The reason for their occurrence is still not clear and is a matter of concern to several investigators. This might be related to the fact that a very small shift in the floating mark position (i.e. in the x-parallax measurement) will result in a quite larger change in the height value due to the very large flying height and the narrow base:height ratio.

iii). In general, the accuracy (RMSE) of a DTM derived from the MC photography will be in the range from $\pm$20m to $\pm$35m, depending on many factors such as the type of terrain, the control for orientation, the skill of the operator, etc. While this will certainly not be sufficient for a high accuracy DTM as required for example for a civil engineering project, it is still quite sufficient for reconnaissance purposes, e.g. for military defence purposes, landscape visualization, aircraft flight simulators and the production of contours for small scale topographic mapping and charting on a regional or national basis.

The discussions carried out in this chapter and the previous chapter (Nine) are purely about practical investigations into the accuracy of the DTMs derived from photogrammetrically measured data. In the next chapter, a theoretical discussion will be carried out to see how reliable are these accuracy figures derived from the experimental test results.

Chapter 11

Effects of Check Points on the Reliability of DTM Accuracy Estimates Obtained from Experimental Tests

Chapter Eleven

Effects of Check Points on the Reliability of DTM Accuracy Estimates Obtained from Experimental Tests

11.1 Introduction

In the previous chapter, some experimental tests have been carried out on the accuracy of digital terrain models derived from photogrammetrically measured data with different sampling intervals for three test areas. In each test, a **set of check points** was used as the "ground truth". Then the points interpolated from the constructed DTM surface were checked against the corresponding check points. After that, the difference of the two heights (DH) at each DTM point was obtained. These errors were used to compute some statistical values such as the mean and standard deviation which were used as a measure of DTM accuracy. In these circumstances, DH was considered as a random variable.

In the case of the experimental tests on DTM accuracy, it is clear that the final DTM accuracy figures estimated from the test results - in this case, the mean and standard deviation values - are definitely affected by the characteristics of the set of check points. To put it in another way, it can be said that the characteristics of the set of check points which were used as the ground truth in the experimental tests have effects on the **reliability** of the final DTM accuracy figures obtained from these tests.

It is obvious that the reliability of the accuracy values estimated from an experimental test is also a problem which is of considerable importance in DTM accuracy tests since, only if the accuracy values are reliable to certain level, can one use the accuracy estimates to evaluate the "goodness" of the digital terrain model which has just been tested. Therefore, this chapter is an attempt to obtain an insight into the effects of the set of check points used in the experimental tests on the reliability of the DTM accuracy values estimated from the test results.

In this chapter, first of all, the concept of reliability in the context of DTM accuracy tests will be introduced and alternative measures for this will be sketched. Then, the effects of the characteristics of a set of check points on the DTM accuracy estimates will be investigated both through a theoretical analysis and by experimental tests.

11.2 Reliability in the context of DTM accuracy tests

Reliability is a concept which is widely used in engineering and industry. It seems pertinent to have a look at how this concept is defined and used in these areas before it can be adopted into the methodology and context of experimental tests on DTM accuracy.

11.2.1 The concept of reliability in engineering and industry

Due to the many differing operational requirements and varying environments existing in engineering and industry, the **concept** of reliability may mean quite different things to different people. Nevertheless, a generally acceptable definition given by the B.S.I. (British Standards Institution) is as follows (Dummer and Winton, 1986):

> "Reliability is the characteristic of an item expressed by the probability that it will perform a required function under stated conditions for a stated period of time."

For example, suppose the life of the bulbs made by a lamp manufacturer is declared to be 1,000hrs (which is the stated period of time required by the above definition), then the reliability is 98% if one tested 100 bulbs of this make and found that 2 of them had shorter lives than declared.

This might belong to one of the simplest examples. In practice, the reliability of an engineering system or structure is much more complicated. However, the detailed discussion of this matter lies outside the interest of this thesis. What is intended here is to adopt the concept of reliability into the context of DTM accuracy estimates.

11.2.2 Reliability in the context of DTM accuracy tests

Obviously, in the context of **experimental tests** on the accuracy of a digital terrain model, there is nothing which is concerned with "a required function under stated conditions for a stated period of time". Instead, what is of concern in this context is "with what probability are the estimated accuracy values (i.e. the mean and standard deviation values) likely to be correct" or "to what degree of accuracy will the accuracy results have been estimated". In any case, it is an obvious fact that the DTM accuracy results obtained from experimental tests are not absolutely certain and one can accept these results only to a certain confidence level. Therefore, the concept of reliability can be adopted into this context since reliability is concerned only with uncertainty.

In some sense, the concept of **reliability** in this context might be

defined as the degree of accuracy to which the DTM accuracy results have been estimated, or as the probability with which the DTM accuracy estimates are correct.

Of course, the reliability of DTM accuracy estimates in the context of experimental tests may be affected by several factors such as the capabilities of the person who has undertaken the work; the program by which the accuracy results have been calculated and recorded; and the characteristics of the set of check points which have been used as ground truth against which the DTM points have been checked.

It is obvious that, the accuracy results will be unreliable if the program which has been used to compute the accuracy figures is not correct. Similarly, if the person whose duty is to carry out the test and to record the accuracy figures is careless, then the accuracy results will also be unreliable or incorrect. However, in this study, it is assumed that these two factors are absolutely reliable. Therefore, only the effect of certain characteristics of the set of check points on the reliability of DTM accuracy estimates will be considered.

A **set of check points** can be characterised by three main parameters, namely, their accuracy; the sample size (i.e. the number of points in the data set); and the location and distribution of the data points. Therefore, the main discussion in this chapter will be about how each of these three main parameters of a set of check points affect the reliability of DTM accuracy estimates in the context of experimental tests.

11.2.3 Alternative measures of reliability

As one can imagine, a **measure** is required for the reliability of DTM accuracy estimates. There may be two types of measure available. One is **qualitative** (or descriptive) and the other type is quantitative (or numerical). For the former, words such as absolutely reliable, most reliable, very reliable, quite reliable, fairly reliable, not so reliable, not reliable, unreliable, most unreliable, absolutely unreliable, etc. can be used. However, in the scientific community, such a statement is not acceptable since the definition of such a term is usually too loose.

For the **quantitative** (or numerical) measures, there are three alternatives as follows:

i). One possible measure is to use the **absolute values** of the accuracy of each of the obtained accuracy estimates, e.g. the value of standard deviation of the obtained standard deviation

estimate. Suppose that the estimated standard deviation values for the DH accuracy is SD(DH), then such an absolute value may denoted as SD(SD(DH)).

ii). Another possible measure is to use a **relative value**, similar to the term "per mil of flying height" which is commonly used to state the accuracy of photogrammetrically measured data. Thus in this case, a percentage value may be quite adequate and thoroughly acceptable. For example, the percentage value of the ratio SD(SD(DH))/SD(DH) might well be adequate.

iii). The third possible way is to use the concept of **"membership"** used in the context of fuzzy sets. Percentage values between 0% and 100% can be used and these values represent the degree of reliability to which an accuracy estimate belongs. In this case, it is not necessary that the percentage value be obtained from the ratio SD(SD(DH))/SD(DH). Instead, the SD(SD(DH)) value is converted into a figure expressing the degree of reliability by a pre-defined function.

There is no fundamental difference between the second and the third approaches. The second value will become the same as the third if the former is stretched into the range of 0% to 100%.

After these introductory discussions and definitions, it is time to look into the matter of the effect of check points on the DTM accuracy estimates.

11.3 Effect of sample size (number) on the reliability of the DTM accuracy estimates

11.3.1 Introduction

The problem of the effect of sample size of the check points on the reliability of DTM accuracy estimates may be considered first. More precisely, the discussion in this context will be about the sample size of the random variable DH. However, in practice, the values for both of them are exactly the same, since the height differences are computed from the check points. Therefore, the number of check points is used as a synonym for the term sample size of the random variable DH.

It seems obvious that the inclusion of more check points in the data set shall lead to a more reliable result. So researchers try to use large sample sizes in order to ensure that the obtained accuracy values will be reliable. For example, in the ISPRS DTM test which was conducted by Commission III's Working Group No.3 (Torlegard et al,

1986), more than 1,800 check points were used in each test area. However, a large number of check points may sometimes be costly to produce and, in some cases, even impossible to provide in the context of DTM accuracy testing.

Therefore, an important question which arises is whether such a large number of check points is necessary. If not, then the obvious follow-up question is what is the minimum number of check points required for a given degree of reliability for the accuracy estimates." That is to say, the important matter in this case is to determine the required minimum sample size (number) for the given degree of reliability required for the accuracy estimates (i.e. the estimated mean and standard deviation values).

Ley (1986) tried to provide a solution to this problem based on his own experience and pointed out that "a sample size of 150 points will guarantee that the subsequent accuracy statement possesses a standard deviation of 10%" (of the estimated standard deviation). This number (150 points) is over 10 times smaller than that used in the ISPRS test. However, he didn't provide any information about how this figure was obtained nor the context in which it occured. Therefore, a theoretical deduction may be both revealing and important.

In an attempt to answer the questions raised above, this section starts with a theoretical analysis; then the theoretical results will be validated with experimental data. The theoretical analysis in this study is based on the assumption that the check points are free of error.

11.3.2 Effect of sample size on the accuracy of the estimated mean value

From statistical theory, it can be found that the **sample size** required for a given degree of accuracy for the accuracy values to be estimated depends on the variation associated with the random variable - DH in the case of the DTM accuracy tests. The smaller the variation, the smaller the sample size that is needed to achieve a given degree of accuracy required for the accuracy estimates. For an extreme example, if the standard deviation (SD) of the height differences was equal to zero, then one check point would be enough no matter how large is the test area or the size of the data set. The required minimum sample size also depends on the given degree of the accuracy requirement itself. A general discussion about the relationship between the sample size, the value of SD and the given degree of the accuracy requirement is given in the following paragraphs.

Let M be the mean of a random sample of size n from a particular

distribution, and u be the true value of the random variable. Then the ratio

$$Y = \frac{M - u}{SD / n^{1/2}} \qquad (11.1)$$

is the **standardized variable** and has approximately the normal distribution N(0,1), even though the underlying distribution is not normal, as long as n (the sample size) is large enough (Hogg and Tanis, 1977).

Suppose the SD of a distribution is known but the value of u (the true value of the random variable) is unknown. Then, for the probability r and for a sufficiently large value of n, a value Z can be found from the statistical table for N(0,1) distribution such that the probability that Y will be within the range from -Z to Z is approximately equal to r; or mathematically

$$P(-Z \leqslant Y \leqslant Z) \simeq r \qquad (11.2)$$

The **closeness** of the approximate probability r to the exact probability depends upon both the underlying distribution and the sample size. When the underlying distribution is unimodal (with only one mode) and continuous, the approximation is usually quite good for even a small value of n (e.g. n=5). If the underlying distribution is "less normal", (i.e. badly skewed or discrete), a large sample size is required to keep a reasonably accurate approximation. However, 20 or 30 is the number which is quite adequate for n in all cases (Hogg and Tanis, 1977).

Substituting Equ.(11.1) into Equ.(11.2) and rearranging it, the following expression can be obtained:

$$P(M - Z\ SD/\ n^{1/2} \leqslant u \leqslant M + Z\ SD/\ n^{1/2}) \simeq r \qquad (11.3)$$

For a given constant S, the percentage of the probability, (100r)%, of the random interval $M \pm S$ including u is called the **confidence interval**, where S is, in fact, the specified degree of accuracy for the mean estimate, M in this case. In general, if the required confidence interval (100r)%=100(1-α)%, then the sample size n can be expressed as the following according to (11.3):

$$S = \frac{Z_r \cdot SD}{n^{1/2}} \qquad (11.4)$$

Where, SD is the standard deviation of the random variable;
S is the given degree of accuracy; and

Z_r is the limit value within which the values of the random variable Y will fall with a probability of r. Its value can be found in the statistical table for the N(0,1) distribution. The mathematical expression is the following:

$$\Phi(Z) = 1 - \alpha/2 \qquad (11.5)$$

and the commonly used values are as follows:
$Z_{r=0.95} = 1.960$; $Z_{r=0.98} = 2.326$; $Z_{r=0.99} = 2.576$.

In the case of a DTM accuracy test, the SD in (11.4) is the expected standard deviation of the final DTM accuracy, and an approximate estimate is required before starting to measure the check points. The value of r is commonly selected as 95%, 98% or even 99%. S is the specified degree of accuracy for the mean - M in this case. Equation (11.4) can also be rewritten as:

$$n = \frac{Z_r^2 \cdot SD^2}{S^2} \qquad (11.6)$$

Where, S is the accuracy requirement for the mean; SD is the expected standard deviation of the DTM and n is the required minimum sample size for the check points with a given confidence level which is expressed by Z_r. The ratio S/SD expresses the proportion of S to SD; thus it is a value which can be used as the reliability of the estimated mean value. If it is denoted as R(M), then Equation (11.6) can be rewritten as follows:

$$n = \frac{Z_r^2}{R^2(M)} \qquad (11.7)$$

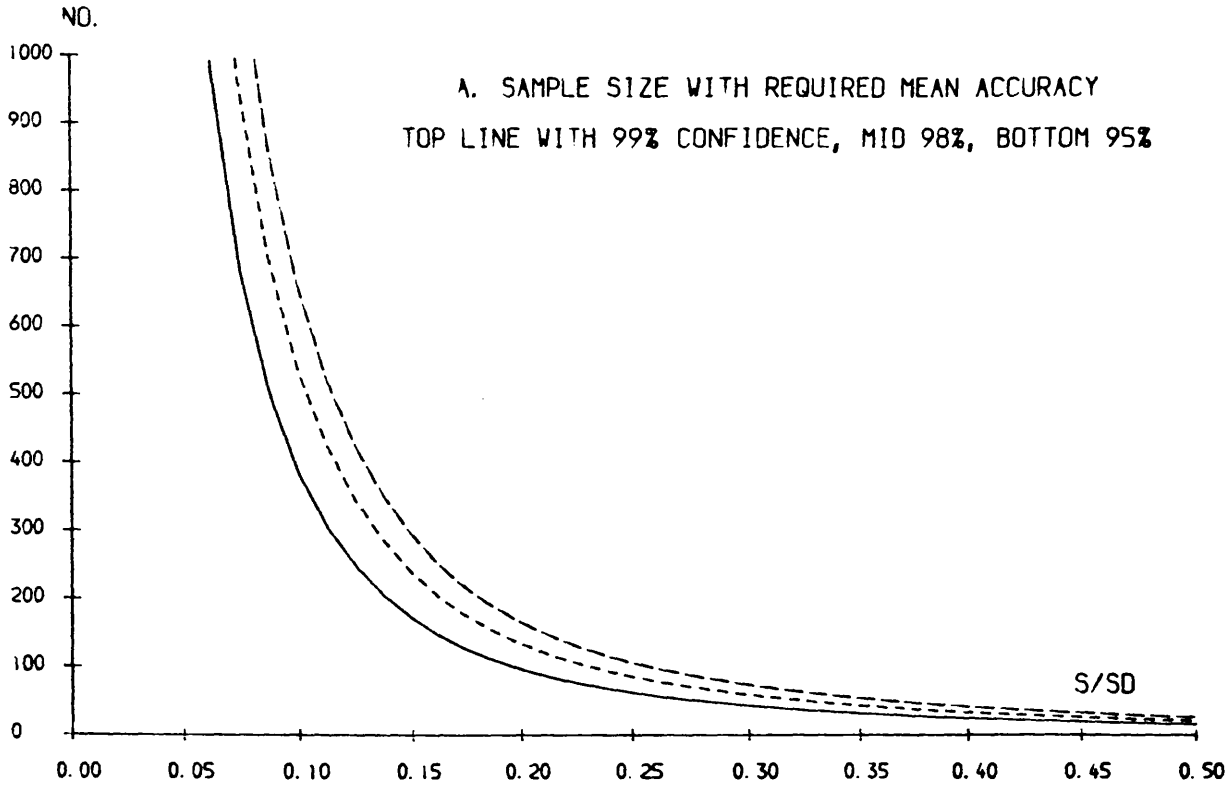
The diagrammatic presentation of Equ.(11.7) is given in Fig.11.1(a). Equ.(11.7) can also be rewritten as follows:

$$R(M) = \frac{Z_r}{n^{1/2}} \qquad (11.8)$$

11.3.3 Effect of sample size on the reliability of estimated SD value

Next, the influence of sample size on the reliability of the SD estimate should be considered. It can be found that the variance of the standard deviation estimated from a sample can be approximately expressed as follows (Burington and May, 1970):

$$VAR(SD(DH)) = VAR(DH) / (2n) \qquad (11.9)$$



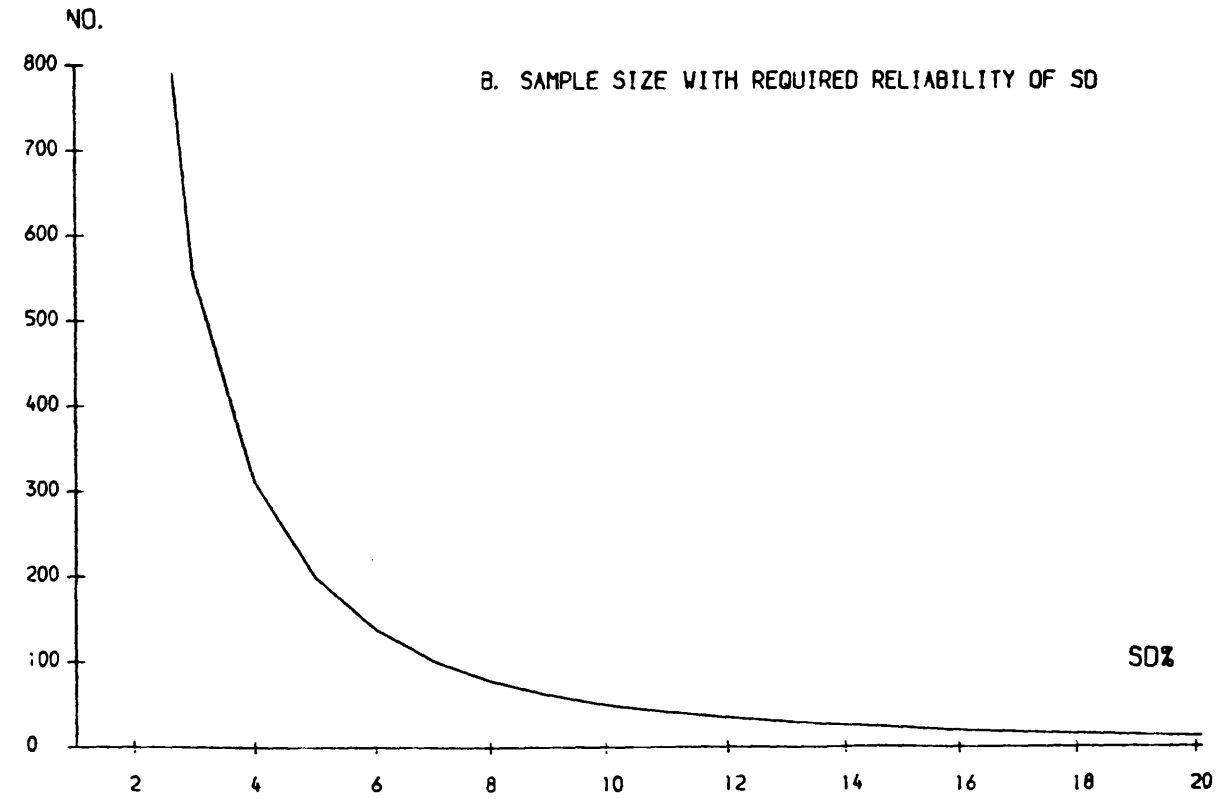


FIG. 11. 1 SAMPLE SIZE WITH REQUIRED ACCURACY

In the context of a DTM set, this would mean that the estimated standard deviation of the DTM errors possesses a precision of $(1/2n)^{1/2}$ times itself if the check points are free of error, or with a precision higher than the critical value which will be discussed later. This can be expressed as a percentage of the estimated variance. It can be written as follows:

$$R(SD) = \frac{1}{(2n)^{1/2}} \times 100\% \qquad (11.10)$$

For example, a sample size of 150, which was given by Ley (1986) as an example, will provides a precision of 6% for the standard deviation estimate. This value is very close to that presented by Ley (1986). To give another example, a sample size of 1,800 will produce a precision of 2% for the standard deviation estimate.

Accordingly, if the reliability requirement for the DTM precision statement is given beforehand, then the required minimum sample size can also be computed from the following:

$$n = \frac{1}{2\,R^2(SD)} \qquad (11.11)$$

Where R(SD) is a percentage value. For example, if R(SD)=10% is the precision required, then from (11.11), it can be computed that the required number for this example is 50. The graphical presentation of (11.11) is shown in Fig.11.1b.

11.3.4 Experimental validation

The discussion given in previous two sections is purely theoretical. One very important question arising from this discussion is whether these criteria can really be applied in practice. To answer this question, some experimental tests are necessary.

This has been done using the data sets which had been generated and used in the ISPRS test. Detailed information about these data sets is given in Chapter 9. Two areas, i.e. Uppland and Sohnstetten, have been selected for this experiment because, from the tests which have been described in Chapter 9, it was found that the occurrence frequencies of large residual errors are very low for these two areas. Thus, the data sets for these two areas are assumed to be very reliable. Therefore, only these two have been used.

The check points were originally arranged in a grid form. For Uppland, the grid size is 69x36 = 2,484 points and 20x104 = 2,080 points for Sohnstetten. However, not every point was measured because a certain number fell in a woodland area or on some other unsuitable feature. In fact, only 2,314 grid nodes were measured for Uppland and 1,892 for the Sohnstetten area. From these points, several sub-sets were selected. These data sets were selected simply by choosing **every nth point** from the data file. The test results are shown in Tables 11.1 and 11.2, where the symbol "$\pm$" before SD and RMSE values is simply omitted.

Table 11.1 shows the variation in the accuracy of the DTM from composite data sets for **Uppland** with the number of check points used. The SD for the Uppland data is obtained from Chapter 9 as $\pm$0.592m. According to (11.6), if the estimated mean should lie within a range of $\pm$0.05m from the true value with 95% confidence, then 535 check points are required for the purpose. However, with the same confidence level, 134 and 273 points will give estimated means within a range of $\pm$0.10m and $\pm$0.07m from the true value, respectively. From the same table, it can also be seen that the results obtained using more than 578 check points are very consistent not only for the mean values (varying within a range of $\pm$0.016m) but also for the SD and RMSE. Below this number, the mean, the SD and the RMSE all show bigger variations. When the number of check points lies within the range between 257 and 578, the mean varies within the range of $\pm$0.065m. When fewer than 115 check points were used, then the figures of these accuracy parameters become very unstable. The results in this table show more or less similar trends to those expressed by Equs.(11.6) and (11.8).

Table 11.2 shows the variation in DTM accuracy with the number of check points for the **Sohnstetten** area. The SD value for the Sohnstetten data set obtained in Chapter 9 is $\pm$0.401m. Also according to (11.6), with 95% confidence, 683, 245, 125, and 62 check points will give the estimated means within the ranges of $\pm$0.03m, $\pm$0.05m, $\pm$0.07m, and $\pm$0.10m from the true value respectively. From Table 11.2, it can be found that the mean varies from 0.153 to 0.173 in a range of 0.02m. At this stage, the SD and RMSE values are very stable.

When the number of check points falls within the range 379 to 237, the mean varies over a greater range of 0.035 (0.154 to 0.189). Also the SD and RMSE values vary over a greater range. When the number of check points lies within the range of 211 to 119, the means vary with a range of 0.063m (0.144 to 0.207m). When the number of check points lies within the range between 106 to 64, the mean varies from 0.085m to 0.236m. It is 0.068m lower and 0.083m higher than the value 0.0153m which is that obtained using all the check points (i.e. 100%) in this test. Accordingly, the RMSE and SD values also vary with a greater range when fewer check points are used. This test again shows that the

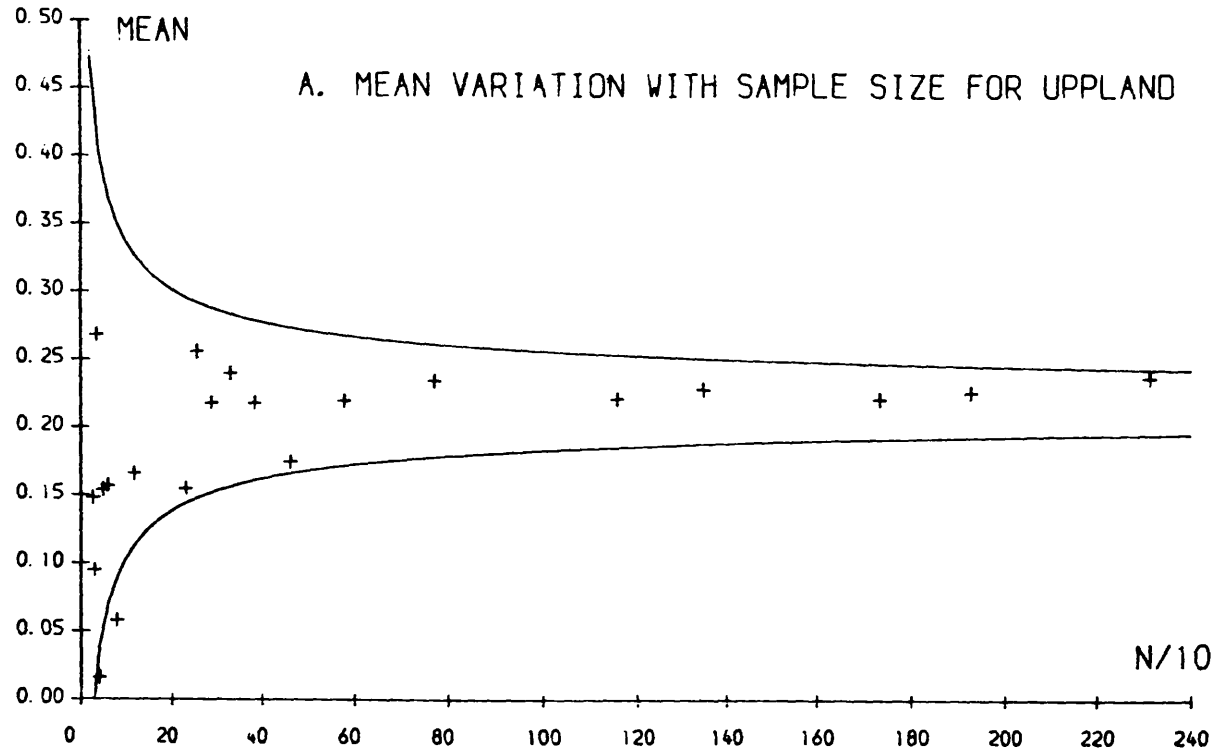
Equs. (11.6) and (11.10) are appropriate.

Table 11.1 Variation of DTM accuracy with number of check points for the Uppland the area

Parameters for check points			Parameters for DTM accuracy		
Fraction	%	Number	RMSE	SD	Mean
1 / 1	100.0 %	2,314	0.636 m	0.590 m	0.238 m
5 / 6	83.3 %	1,928	0.618 m	0.575 m	0.227 m
3 / 4	75.0 %	1,735	0.614 m	0.573 m	0.222 m
7 /12	58.3 %	1,349	0.618 m	0.574 m	0.229 m
1 / 2	50.0 %	1,157	0.615 m	0.574 m	0.222 m
1 / 3	33.3 %	771	0.622 m	0.576 m	0.235 m
1 / 4	25.0 %	578	0.612 m	0.571 m	0.220 m
1 / 5	20.0 %	462	0.592 m	0.566 m	0.175 m
1 / 6	16.7 %	385	0.619 m	0.579 m	0.218 m
1 / 7	14.3 %	330	0.597 m	0.547 m	0.240 m
1 / 8	12.5 %	289	0.587 m	0.545 m	0.218 m
1 / 9	11.1 %	257	0.623 m	0.568 m	0.256 m
1 /10	10.0 %	231	0.606 m	0.586 m	0.155 m
1 /20	5.0 %	115	0.566 m	0.545 m	0.166 m
1 /30	3.3 %	77	0.574 m	0.571 m	0.058 m
1 /40	2.5 %	58	0.570 m	0.548 m	0.157 m
1 /50	2.0 %	47	0.759 m	0.743 m	0.154 m
1 /60	1.7 %	39	0.580 m	0.580 m	0.016 m
1 /70	1.4 %	34	0.437 m	0.346 m	0.268 m
1 /80	1.25%	29	0.554 m	0.546 m	0.095 m
1 /90	1.1 %	26	0.630 m	0.613 m	0.148 m

Table 11.2 Variation of DTM accuracy with number of check points for the Sohnstetten Area

Parameters for check points			Parameters for DTM accuracy		
Fraction	%	Number	RMSE	SD	Mean
1 / 1	100.0 %	1,892	0.429 m	0.401 m	-0.153m
5 / 6	83.3 %	1,580	0.422 m	0.391 m	-0.159m
3 / 4	75.0 %	1,419	0.429 m	0.398 m	-0.160m
7 /12	58.3 %	1,104	0.430 m	0.393 m	-0.173m
1 / 2	50.0 %	946	0.423 m	0.395 m	-0.152m
1 / 3	33.3 %	631	0.421 m	0.385 m	-0.170m
1 / 4	25.0 %	473	0.442 m	0.405 m	-0.178m
1 / 5	20.0 %	379	0.443 m	0.405 m	-0.159m
1 / 6	16.7 %	316	0.409 m	0.377 m	-0.159m
1 / 7	14.3 %	271	0.437 m	0.401 m	-0.172m
1 / 8	12.5 %	237	0.439 m	0.396 m	-0.189m
1 / 9	11.1 %	211	0.415 m	0.389 m	-0.144m
1 /10	10.0 %	190	0.417 m	0.372 m	-0.189m
1 /12	8.3 %	158	0.406 m	0.363 m	-0.181m
1 /14	7.1 %	136	0.449 m	0.401 m	-0.201m
1 /16	6.25%	119	0.460 m	0.410 m	-0.207m
1 /18	5.56%	106	0.398 m	0.379 m	-0.122m
1 /20	5.0 %	95	0.477 m	0.414 m	-0.236m
1 /25	4.0 %	76	0.371 m	0.361 m	-0.085m
1 /30	3.3 %	64	0.349 m	0.291 m	-0.193m
1 /40	2.5 %	48	0.518 m	0.420 m	-0.302m
1 /50	2.0 %	38	0.346 m	0.315 m	-0.143m
1 /60	1.7 %	32	0.358 m	0.298 m	-0.197m
1 /70	1.4 %	28	0.492 m	0.410 m	-0.272m



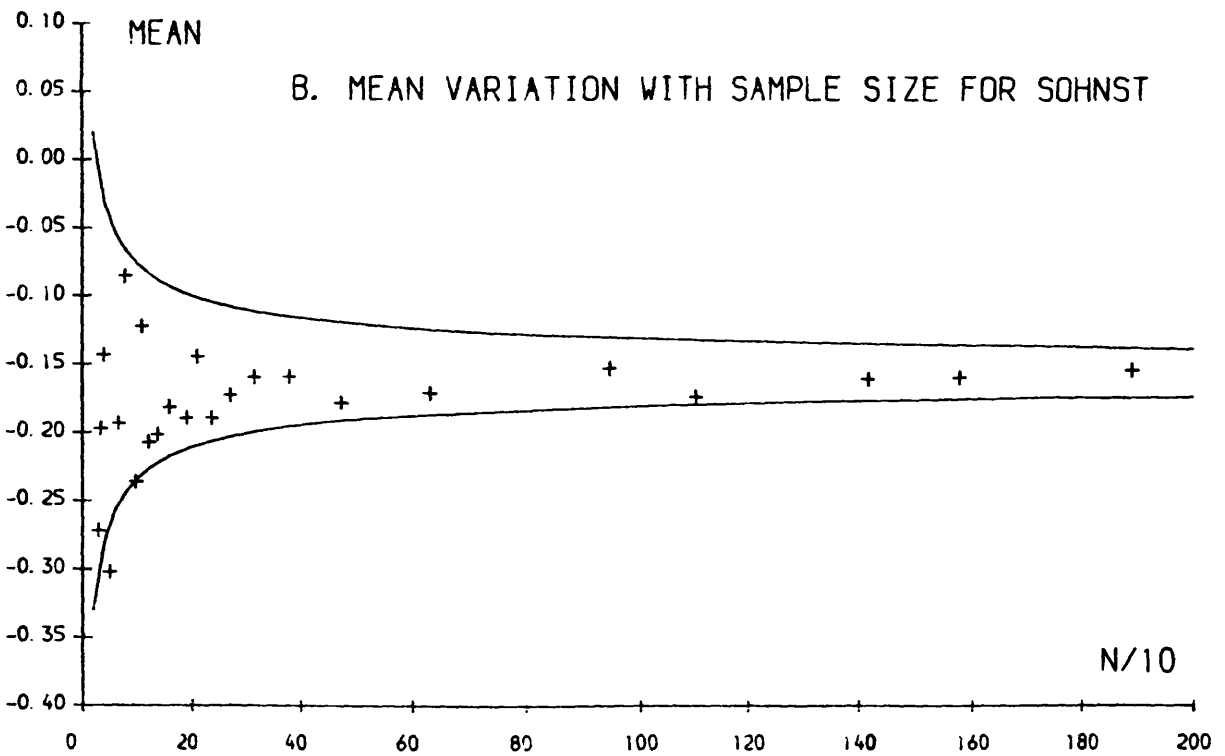
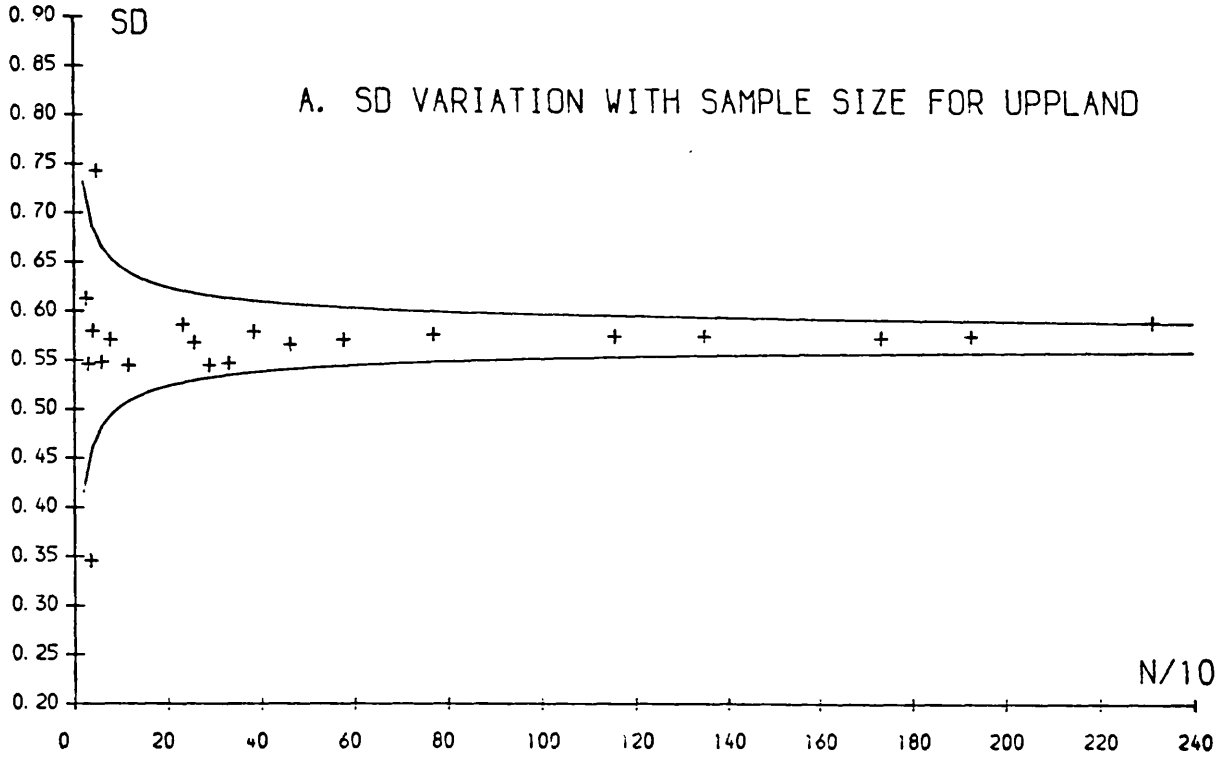


FIG. 11.2 VARIATION OF MEANS WITH SAMPLE SIZE



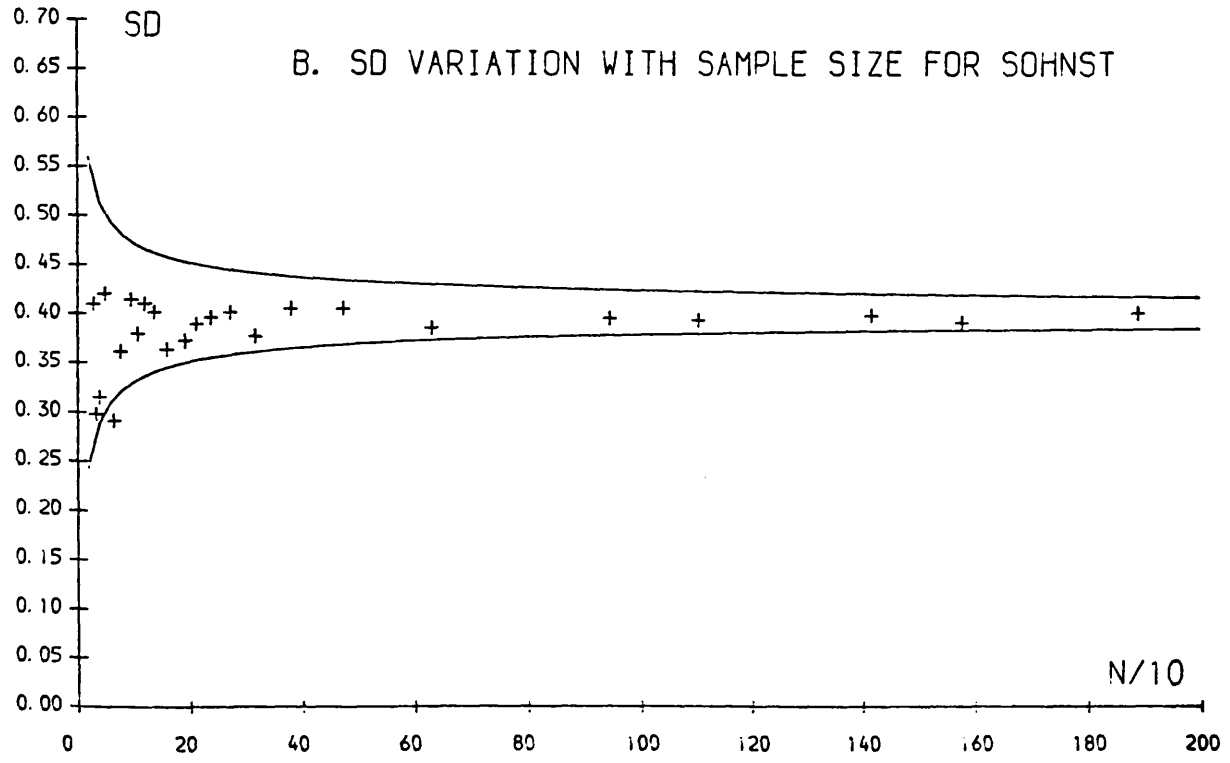


FIG. 11.3 VARIATION OF SD WITH SAMPLE SIZE

More intuitively, these data values are presented graphically in Fig.11.2 and Fig.11.3. The continuous lines represent the variation ranges which are predicted from purely theoretical considerations. The lines in Fig.11.2 are produced according to Equ.(11.4), where a 95% confidence level is selected; and 0.220m and -0.155m are used as the "true" values of the means for the Uppland and Sohnstetten areas respectively. Here it needs to be pointed out that the term S used in Equ.(11.4) is the given degree of absolute accuracy but not the precision. The latter is well-known to topographic scientists as follows:

$$VAR(M) = VAR(DH) / n \qquad (11.12)$$

The lines in Fig.11.3 are produced according to Equ.(11.10); using 0.575 and 0.401 as the SD values for Uppland and Sohnstetten areas respectively. Symbol "+" presents the points obtained from experimental tests which have been listed in Table 11.1 and Table 11.2. These diagrams appear to prove the validity of the theoretical discussions which were conducted in the previous two sections, at least in the context of the ISPRS test data.

11.4 Effects of errors in the check points on the reliability of the DTM accuracy estimates

11.4.1 Introduction

Next the matter of the errors in the check points which were used as ground truth in the experimental tests and their effects on the reliability of the DTM accuracy estimates requires discussion.

As stated before, the sample of DH values is obtained by comparing the heights of the points given by the DTM with the values given by the check point data. The errors present in the check points themselves affect the values of DH, and thus the DTM accuracy estimates. It is obvious that the smaller the errors in the elevation values of the check points, the less influence they will have on the estimated DTM accuracy figures. Thus eventually, if the check points are free of error, they have no effect, i.e. they will not contribute to the DTM accuracy estimates. In practice, such a case can never occur. Therefore, the important thing is to establish the relatioship between the accuracy of the check points and the reliability of the final DTM accuracy estimates so that the maximum tolerable accuracy for the check points can be determined for a given degree of reliability required for the final DTM accuracy figures to be estimated.

11.4.2 Accuracy requirement for the check points

In the context of topographic science, in most cases, the accuracy of the check points is specified in terms of root mean square error (RMSE). In the present discussion, this RMSE value is assumed to be the same as the standard deviation (SD). Therefore, in this context, the important thing is to find the relationship between the SD value of the check points and the given degree of reliability required for the final standard deviation estimate which depends on the sample size of the check points.

Let DH_2 be the error involved in the check points and DH_1 be the true height difference. Then the overall error (DH) is as follows:

$$DH = DH_1 + DH_2 \qquad (11.13)$$

By applying the error propagation law to Equation (11.13), the following expression can be obtained:

$$VAR(DH) = VAR(DH_1) + VAR(DH_2) \qquad (11.14)$$

The variance of DH could be the sum of a few random variables. In this study, it is split into two, i.e. $VAR(DH_1)$ and $VAR(DH_2)$. The value of VAR(DH) itself is not of interest, but the value of $VAR(DH_1)$ is. An attempt may be made to estimate the latter through the former because only the former can be known. The attempt which is made here is to find a critical value for $VAR(DH_2)$ so that the value of VAR(DH) is still acceptable as being representative of $VAR(DH_1)$.

Also as quoted in Section 11.3.2, the standard deviation estimated from a sample of size n has a variance as follows:

$$VAR(\ SD(DH_1)\ ) = VAR(DH_1) \ / \ (2n) \qquad (11.15)$$

Therefore, the acceptable range for SD(DH) to deviate from $SD(DH_1)$ can be expressed as follows:

$$SD(DH_1) - SD(DH_1)/(2n)^{1/2} < SD(DH) < SD(DH_1) + SD(SD(DH_1)/(2n)^{1/2} \qquad (11.16)$$

It is much more convenient to use a single value, so the square root of these two terms is used as the representative value since they are independent. Then the following equation can be obtained:

$$\begin{aligned} VAR(DH) &= VAR(DH_1) + VAR(DH_1) \ / \ (2n) \\ &= (2n+1)\ VAR(DH_1) \ / \ (2n) \end{aligned} \qquad (11.17)$$

Combining Equs.(11.17) and (11.14) with a simplification, the following

expression can be derived:

$$VAR(DH_2) = VAR(DH) / (2n+1) \quad (11.18)$$

It is more convenient to express this criterion in terms of the standard deviation. So Equation (11.18) can be converted to the following form:

$$SD(DH_2) = SD(DH) / (2n+1)^{1/2} \quad (11.19)$$

Let $K = SD(DH_2)/SD(DH)$, then equation (11.19) can be rewritten as follows:

$$K = 1 / (2n + 1)^{1/2} \quad (11.20)$$

Where, K is a function of the sample size, n. A graphic presentation is shown in Fig.11.4. For a given sample size which is determined by the reliability requirement discussed in the previous sections (11.3.2 and 11.3.3), the critical value for the required accuracy of the check points can be determined by equation (11.19). In which case, $SD(DH_2)$ may be given a special annotation, thus denoted as SD(t) in this context.

Obviously, the value of K decreases with an increase in n. This means that, with the increase in n, the precision of the estimated $SD(DH_1)$ value increases. And the higher the precision of $SD(DH_1)$, the greater the influence of the check points with the same precision on the reliability of the estimated $SD(DH_1)$ which is approximated by SD(DH).

As discussed before, if the precision of the check points is higher than the critical value, then SD(DH) can be used to approximate $SD(DH_1)$ since the former is still within the precision range of the latter. The reliability of SD(DH) can still be approximated by Equation (11.10).

On the other hand, if the standard deviation of the check points is larger than the value of SD(t), then the estimated value of SD(DH) is not as reliable as it should be in theory with the same sample size. Alternatively, it can said that the value of SD(DH) possesses a larger variance than the theoretical value for that sample size. Therefore, the value of SD(DH) is not reliable enough to be used to represent $SD(DH_1)$. In this case, some special treatment needs to be given to SD(DH).

The traditional method is to subtract a value of $SD(DH_2)$ from SD(DH) according to the principle of error propagation expressed by Equation (11.14). However, two problems then arise. The first concerns the reliability of such a treatment. The second concerns the limits of $SD(DH_2)$ within which such an operation can be carried out and still be

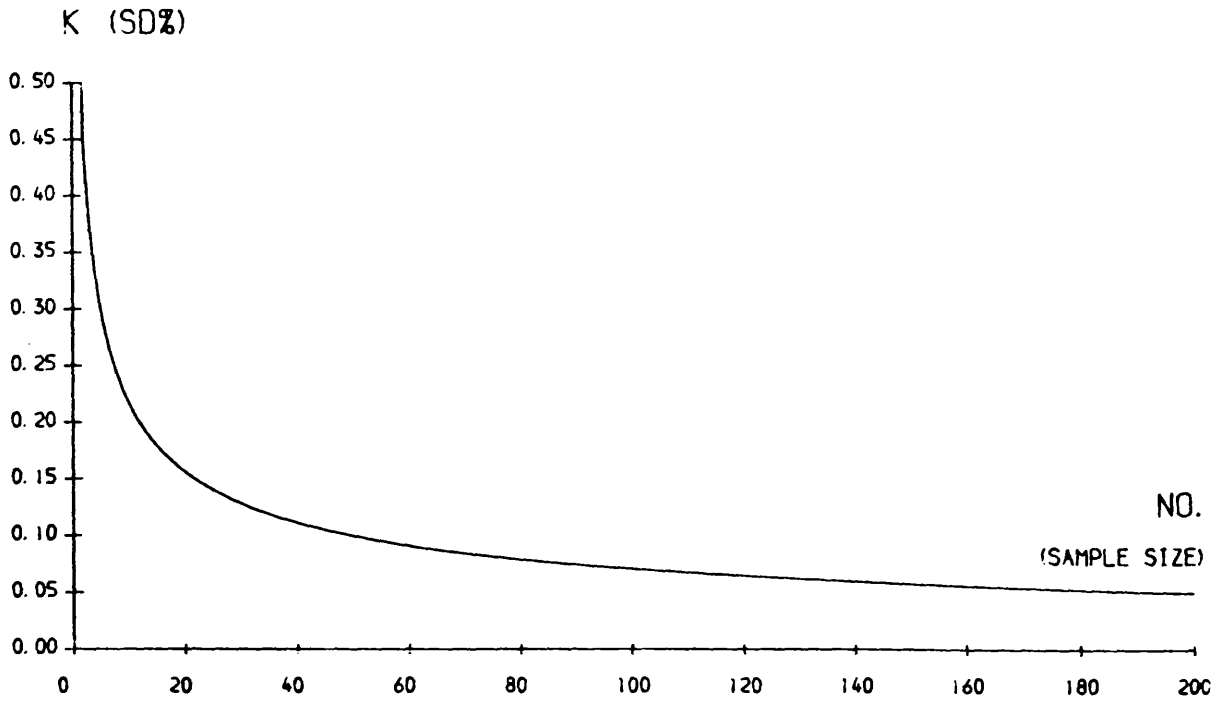


FIG. 11.4. REQUIRED ACCURACY OF CHECK POINTS IN TERMS OF SD% WITH SAMPLE SIZE

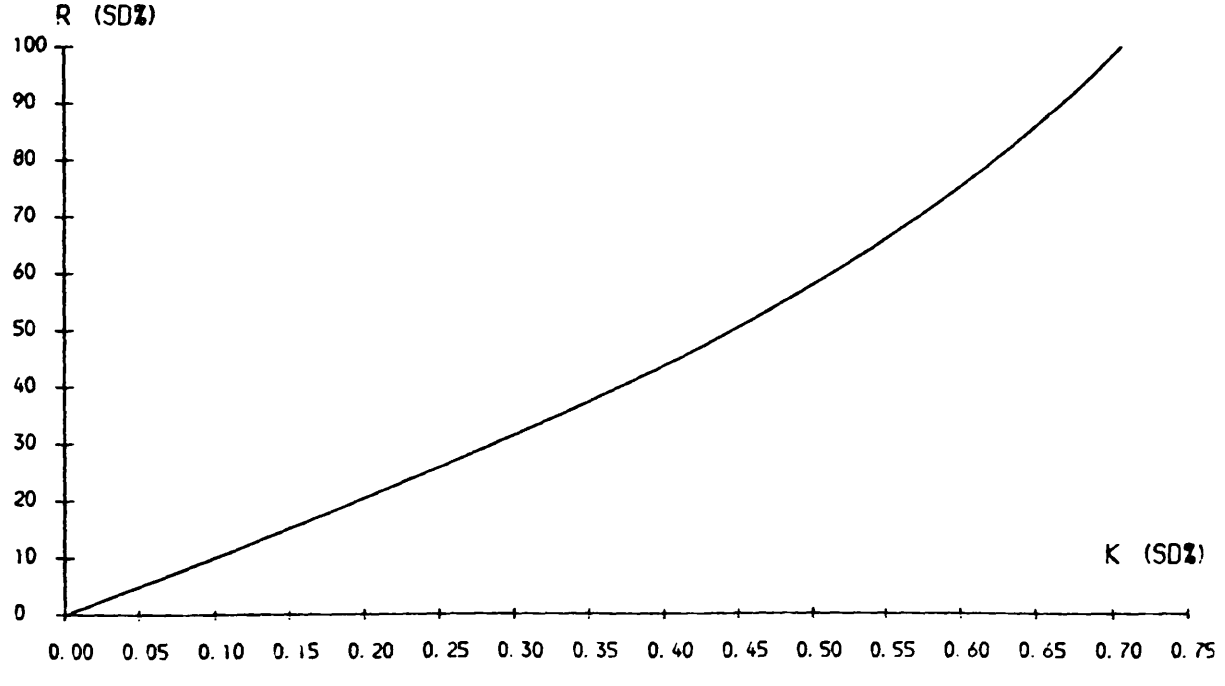


FIG. 11.5 RELIABILITY OF ESTIMATED SD WITH THE ACCURACY OF THE CHECK POINTS

valid. In the extreme case, it is still valid to carry out such an operation even if $SD(DH_2)$ is greater than or equal to $SD(DH_1)$. These are problems which are difficult to solve since there appears to be no theory to offer guidance on this matter. Therefore, the criterion expressed by Equation (11.19) might be used when the DTM accuracy test experiment is carried out.

11.4.3 Accuracy of check points and the reliability of the standard deviation estimate

From the discussions conducted in the previous section, it can be concluded that if the standard deviation of the check points is smaller than the critical value set out in this section, then their effect is negligible. However, if the check points have a standard deviation larger than the critical value, then the precision of check points itself affects the reliability of the estimated accuracy figures. This section will discuss these effects.

Substituting Equ.(11.20) into Equ.(11.10), the following relationship can be obtained:

$$R(SD) = \frac{K}{(1 - K^2)^{1/2}} \times 100\% \qquad (11.21)$$

This formula expresses the relationship between K (the standard deviation of the check points in terms of the percentage of the standard deviation of the final DTM) and the reliability of the precision estimates. For example, if K=0.09, then R=9.0%. The graphical representation of Equ.(11.21) is shown in Fig.11.5.

The reliability of the estimated standard deviation figure R(SD) derived from both Equ.(11.10) and Equ.(11.21) should be very similar if the accuracy (precision) is higher than the critical value. However, when the accuracy of the check points is lower than the criterion which has been set, then the value of reliability computed from Equ.(11.21) will be much lower than that from Equ.(11.10). Equation (11.21) also shows that, if the SD of the check points is 70.7% of the DTM SD, then the SD of the estimated standard deviation - SD(SD(DH)) - will be equal to the SD(DH) itself. This confirms what has been stated before - namely that the accuracy of the check points affects the reliability of the precision estimates if it is lower than the critical value set by the formula given in Equ.(11.19).

11.5 Effect of the distribution of check points on the reliability of the accuracy estimate

Another important concern with the check points used for the DTM accuracy test is their distribution.

11.5.1 Introduction

The distribution of the check points can be characterised by their locations and patterns. In the ISPRS test, the check points are in a grid pattern. The question must be raised as to whether such a pattern is suitable. If not, then it poses the question as to what kind of distribution is desirable. Ley (1986) has made some efforts to answer this question. He stated that "an accuracy assessment of a DTM should be based on a sample of heights taken from the entire model". He also points out that such "a sample of points should include both the recorded (measured) and interpolated heights." However, the answer to the question as to their distribution is still not complete.

Therefore, it is of interest to know how this factor affects the reliability of the DTM accuracy figures to be estimated from test results. If this was known, then the desirable distribution of check points could then be determined.

In this section, an attempt will be made to discuss this particular matter from the viewpoint of statistical theory. An experimetal test has also been carried out to see if such a theoretical analysis is applicable to DTM practice.

11.5.2 Theoretical discussion

A serious shortcoming of using check points located in a grid pattern is that they then represent a **systematic** sample. In this case, if the first point is sampled, then the positions (locations) of all other points are definitely determined. Such a sample is evenly distributed whereas the procedure which has been discussed for use in an DTM accuracy assessment is based on random sampling. From this point of view, a gridded data pattern is not so appropriate. Thus from the purely theoretical standpoint, in order to make such a statistical procedure applicable, random sampling is desirable.

The use of a grid pattern for check points may be the result of the thought that the DH values in some parts of the area being tested may be greater than those in other parts and, that the sample is representative only if the points are so distributed. But such a line of thinking would ignore the pre-requisite for such a statistical test - namely that the sample should come from the same distribution - because

the fact that the DH values in some parts of the test are greater than those in other parts means that they are not from the same sample space or population. If the stated pre-requisite should be applied, then the large values of DH should also be randomly distributed. Therefore, the use of a gridded pattern of check points is not always sound. The advantages of using it are, in the first place, its convenience and efficiency in terms of implementing a sampling and measuring strategy in a stereo-plotting machine, and its convenience in terms of the resulting data structure which can be implemented in the computer used for the processing of the data.

In this case, the concept of **random sampling** is very clear. It means that there is no intention to select a point in a specific position (location) so that any point, including the recorded points, has an equal chance of being measured at every time of sampling.

Finally, it should be noted that the remarks made in this section are based solely on a purely theoretical analysis and may be not so suitable in practice since the terrain surface is certainly not the result of a purely stochastic process.

11.5.3 Experimental test

The two ISPRS test areas, Uppland and Sohnstetten, have again been used for this purpose. The aim of this test is to find how DTM accuracy estimates vary with different distributions of check points.

The first step in this test is to select randomly some sets of check points with a certain size (number) from the original data sets (1,892 points for Sohnstetten area and 2,314 for Uppland). In this test, for each area, 15 sets of check points have been used, each with a sample size of 500 points. The randomness of the selection was achieved by using a set of random numbers from a uniform distribution which are generated by computer using a NAG routine. In generating the random numbers, the range is determined by the total number of points in the original data set. For example, for the Uppland area, the random numbers lie within 1 to 2,314. After this, those check points with the same numbering as the generated random numbers are taken from the data set and form the sample.

The test results are listed in Table 11.3. Some standard statistical parameters compiled from these results are given in Tables 11.4 and 11.5. In the computation of the percentage values, the arithmetic means are assumed to be the best estimates of these values. The expected tolerable values are computed according to the theoretical formulae set out in the previous sections.

Table 11.3 Accuracy results for the randomly selected check points for the Uppland and Sohnstetten areas (No.=500)

File No.	Results for Uppland Area			Results for Sohnst Area		
	RMSE (±m)	SD (±m)	Mean (m)	RMSE (±m)	SD (±m)	Mean (m)
No1	0.628	0.589	0.219	0.426	0.397	-0.155
No2	0.589	0.543	0.228	0.421	0.390	-0.158
No3	0.603	0.565	0.204	0.417	0.387	-0.157
No4	0.648	0.585	0.278	0.412	0.382	-0.155
No5	0.633	0.595	0.216	0.403	0.369	-0.160
No6	0.621	0.586	0.207	0.407	0.384	-0.136
No7	0.637	0.597	0.224	0.437	0.415	-0.138
No8	0.601	0.565	0.205	0.448	0.425	-0.143
No9	0.629	0.570	0.227	0.453	0.427	-0.152
No10	0.630	0.594	0.212	0.398	0.372	-0.142
No11	0.637	0.593	0.232	0.451	0.426	-0.149
No12	0.623	0.570	0.252	0.431	0.389	-0.186
No13	0.622	0.586	0.208	0.417	0.385	-0.162
No14	0.618	0.567	0.249	0.424	0.397	-0.148
No15	0.617	0.562	0.235	0.408	0.381	-0.145

Table 11.4 Statistical estimates for the SD

		Uppland Area	Sohnst Area
Average Value (AV)		0.5778	0.395
SD of Distribution		0.015	0.019
SD/AV	Computed	2.64%	4.72%
	Expected	3.16%	3.16%

From these results, it can be seen that the precision of the standard deviation for the Uppland data set behaves very well, but that for the Sohnstetten data set is much larger than expected. Using another

measure - the mean, all the values derived from both the Uppland data set and the Sohnstetten data set fall within the range expected.

Table 11.5 Statistical estimates for the mean

		Uppland Area	Sohnst Area
AV	(Average Value)	0.226	-0.152
SD	Computed	0.020	0.012
	Expected	0.026	0.018
S	Max. (computed)	0.052	0.034
	99% Confidence	0.067	0.046
	98% Confidence	0.060	0.041
	95% Confidence	0.051	0.035

11.5.4 Discussion of the test results

Of course, the variation in the accuracy results may also be related to the roughness and/or the steepness of the terrain surface. The fact that the results for Uppland behave better could be due to the smaller slope angles which prevail in the area. The results could also have been affected by the errors in the check points themselves. This test shows that, to a certain extent, random sampling over the entire area (without taking into account geographical location) is a method which is acceptable for the creation and acquisition of check points.

Before ending this discussion, some remarks on the randomness of the check points used in this test need to be made. In this experimental test, nominally, the check points were randomly sampled with a size of 150 points from the entire set of check points. However, in practice, truly random numbers can only be obtained by rolling dice, dealing cards or be generated by special mechanical machines. The randomness which were achieved by the NAG routine is always doubtful since the random numbers generated by computer software follow certain rules specified by algorithms. The following is an example:

The uniform random numbers X_1, X_2, X_3, ... which were used in this test may be generated by a recurring arithmetic process as follows (Frodesen et al, 1979):

$$X_{i+1} = g(X_i, X_{i-1}, \ldots, X_{i-k}) \qquad (11.22)$$

where g is some generating function and k is a constant usually given as 1 or 2. In this case, each X_{i+1} in the sequence will therefore be

completely determined by its predecessors and the given starting values. Thus the sequence is not really random, but it will appear to be so and the results are suitable for most practical applications. The sequence will always be periodic, with a cycle of certain numbers. The sequence can be repeated endlessly if one likes. The random numbers generated in such a way are usually called **pseudo-random** or quasi-random numbers.

11.6 Discussion and conclusion

All the theoretical discussion set out in the previous sections of this chapter are based purely on statistical theory - that is to say, they are valid in theory. The theory on sample size has been confirmed by the limited tests in this study. No tests on the theory related to the accuracy of check points have been carried out in this study because it is too difficult to acquire a set of samples of check points with different accuracies. On the random sampling of check points, one test shows quite encouraging results, but the other is not so satisfactory. Therefore, more studies on the applicability of these theories to the practice of DTM accuracy test still need to be carried out because practical situations are not as perfect as those required by statistical theories.

The standard deviation of the DTM is assumed to be known beforehand both for the estimation of required minimum sample size and the required accuracy (especially the precision) of check points. It needs, therefore, to be estimated at first. However, a rough estimate may serve the purpose.

From the preceding discussion, some conclusions might be drawn for the check points used in the experimental tests on DTM accuracy as follows:

a). The accuracy and the reliability of the final DTM accuracy estimates are affected by the sample size (the number) of the check points used in the experimental tests. In a reverse way, it can be said that the **required minimum sample size** is determined by the given degree of accuracy or the reliability requirement. A general guide to the required values can be derived from Equations (11.6) and/or Formula (11.10), if the check points are free of errors.

b). The reliability of the estimated DTM accuracy figures are also affected by the accuracy of check points. Again, the **accuracy** of check points required for a given degree of reliability can also be determined by Equ.(11.19).

c). The **reliabibility of the estimated standard deviation figure** was expressed in terms of percentage of the estimated value. It can be obtained through the use of Equs.(11.8) and (11.21). When the precision of the check points is lower than the critical value, then Equ.(11.21) should be used to compute the reliability factor.

d). The check points should be **sampled** randomly from the entire testing area (and preferabably as a result of a very even distribution). In this context, the use of the word "randomly" is meant to convey the concept that every point, including the recorded points, has the same chance of being selected every time sampling is carried out.

e). Only if the sample size is increased and the accuracy of the check points is improved at the same time, can the **reliability** of the final estimates be improved. It may be very difficult to implement the second of these criteria.

The discussion carried out in this chapter has been concerned with the effects of the check points which were used in the experimental tests on DTM accuracy, on the reliability of the final DTM accuracy estimates obtained from the tests. In the next chapter, an experimental test on the accuracy of DTMs derived from space photographs will be presented.

Chapter 12

Mathematical Models of the Accuracy of Digital Terrain Model Surfaces

Chapter Twelve

Mathematical Models of the Accuracy of Digital Terrain Model Surfaces

12.1 Introduction

After the final DTM surface has been constructed, its fidelity needs to be validated. In the context of digital terrain modelling, the validation of the fidelity of the constructed DTM surface is usually referred to as the **DTM accuracy assessment.**

DTM accuracy assessment can be carried out in different ways, i.e. either through theoretical analysis or via an experimental test, or a combination of them. As one can imagine, the feasibility of establishing a mathematical model for DTM accuracy through experimental tests has some limitations since the results can only be obtained from special cases. Furthermore, in doing so, an extensive series of tests need to be carried out. This would be very costly and time-consuming, and it might even be impossible to execute in some cases. Therefore, a theoretical analysis is desirable.

In respect of a theoretical analysis of DTM accuracy, as mentioned in the introductory chapter, quite a lot of efforts have been already made by several investigators such as Makarovic (1972), Kubik and Botman (1976), Frederiksen (1980), Tempfli (1980), Frederiksen et al (1986). However, as will be shown from the evaluation of these models which will be given later in this chapter, each of them has certain difficulties or undesirable features. Therefore, the construction of a more comprehensive model is desirable. Indeed, this is the main concern of this chapter.

In this chapter, first of all, the approaches which might be used for the theoretical analysis of DTM accuracy will be discussed; then some of the existing DTM accuracy models will be evaluated; after that, new mathematical models will be established; and finally, these models will be validated using the experimental test results which have been presented in Chapter 9 and evaluated using the seven standards for judgement of such models given in Chapter 2.

12.2 Approaches for DTM accuracy assessment

Unavoidably, some errors will be present in each of the three dimensions of the spatial (X,Y,Z) coordinates of the points occuring on digital terrain model surfaces. Two of these (X and Y) are combined to give the **planimetric error** while the third is in the vertical (Z) direction and is referred to as **the elevation error.** The planimetric error is also known as the **horizontal error,** and the elevation error as

the **vertical error.**

The assessment of DTM accuracy can be carried out in two different modes, by which either the planimetric accuracy and the elevation accuracy can be assessed separately or both are assessed simultaneously. For the former, accuracy results for the planimetry can obtained separately from the accuracy results in the vertical direction. However, for the latter, an accuracy measure for both error components together is required.

Ley (1986) compared the relief portrayal in conventional map production with that resulting from DTM production, and discussed some methods for assessing the accuracy or fidelity of DTMs using the separate assessment approach. He discussed four possible **approaches** for assessing the **elevation accuracy** of DTMs, namely, (a) that predicted by production; (b) that predicted by area; (c) evaluation by cartometric testing; and (d) evaluation by means of diagnostic points. Among them, the first two can be used for theoretical analysis.

(a). One method for implementing the first approach (i.e. that **predicted by production**) is to "assess the likely errors introduced at the various production stages together with an assessment of the vertical accuracy of the source material" (Ley, 1986). Then the accuracy of the final DTM is the consequence or concatenation of the errors involved in all these stages.

(b). The second approach (the elevation accuracy **predicted by area**) is based on the fact that the vertical accuracy of contour lines on topographic maps is highly correlated with the mean slope of the area. There might be an uncertainty about whether it is also the case in the context of DTMs. In fact, the testing carried out in the British Military Survey showed a positive answer (Ley, 1986).

(c) and (d). The third and the fourth approaches are actually not suitable for a theoretical analysis. Instead, they describe methods of experimental investigation. The third approach is the most commonly used one in practice and has also been used in this project (see Chapter 9).

Ley (1986) also mentioned three approaches for the assessment of the **planimetric accuracy** of DTMs, namely, (i) no error; (ii) a predictive approach; and (iii) through height. However, as he also mentioned, it is difficult to bring these into practical use. And it seems to the author that no better methods have ever been proposed so far. Therefore, the assessment of the planimetric accuracy of a DTM will not be discussed in this thesis.

The reason for undertaking the assessment of DTM accuracy combining both the horizontal and vertical directions has been stated by Ley (1986) as follows:

i). Terrain is composed of an infinite number of points located on a three-dimensional surface. The uniqueness of each data point and its relationships with surrounding points are destroyed by dividing the matrix of three-dimensional coordinates into two distributions.

ii). By undertaking the vertical and horizontal tests separately, there is a danger of including some error components twice.

The simultaneous assessment of DTM accuracy requires a measure which is capable of characterising the accuracy in three dimensions. Based on his research, Ley (1986) suggested that a comparative measure of the mean slopes between the DTM surface and the original terrain surface may be appropriate. However, such a measure may be not quite acceptable, at least at present, since people have got used to the measures for the errors being expressed in 3-D coordinates. Therefore, the feasibility of applying this method to DTM accuracy assessment still needs to be investigated. Unfortunately, recent experimental tests carried out by Ley (1990, personal communication) using this approach also show very disappointing results. Therefore, the assessment of DTM accuracy in such a mode will also not be discussed in this thesis.

For the reasons given above, the assessment of vertical accuracy in this study will be carried out only through theoretical analysis. Actually, the development of mathematical models of DTM accuracy in this study is based on the combination of first two approaches (for vertical accuracy assessment) described by Ley (1986).

12.3 Evaluation of existing mathematical models

The work of establishing mathematical models for the height accuracy of digital terrain models started in the early 1970s. Makarovic (1972) in the ITC of the Netherlands did pioneering work in this respect. After this, a number of efforts have been made by several investigators using a variety of mathematical tools such as Fourier transformation, statistics, regionalized variable theory (geo-statistics), etc. By means of these tools, some mathematical models have been established for the prediction of DTM accuracy. These existing models will be evaluated both by theoretical analysis and through experimental tests in this section.

12.3.1 Theoretical evaluation of existing mathematical models

Makarovic (1972) used the mathematical tool of **Fourier analysis** to investigate the fidelity of the DTM surface. He considered the data sampling and the reconstruction from sinusoidal functions. The fidelity of the reconstructed surface is represented by the ratio of the mean value of the magnitude of the linearly constructed sinusoidal waves to the amplitude of the input waves. **Transfer functions** can also be derived for different interpolation techniques. Makarovic (1974) then tried to convert the fidelity figures into standard deviation values. Later, Tempfli and Makarovic (1979) made an intensive study of the transfer functions for a number of interpolation functions. In this way, the accuracies of different digital terrain models can also be compared for different types of terrain surface. As Ackermann (1979) has commented, "In principle, this theory is complete. If the frequency distribution of a terrain is known, all questions regarding point density, interpolation method and accuracy can be answered according to Makarovic. The task remains to investigate the frequency distribution of different terrain types and to relate the corresponding theoretical and empirical accuracy results". Unfortunately, this task still largely remains to be fullfiled up to the present time.

Tempfli (1980) found that the knowledge gained only about the transfer function is not sufficient if the required quality of a DTM is to be specified by a standard deviation value or if the law of error propagation should be applied. By considering the digital terrain modelling system as a linear system, Tempfli (1980) then tried to estimate the accuracy of DTM by a **spectral analysis** of such a linear system. As in the case of Makarovic's theoretical model, the investigation into the frequency distribution of different terrain types must be carried out before any use of this model can be made. Also, experimental tests should be carried out to see how it works in practice. Unfortunately, once again, such work still remains to be carried out up to the present time.

Kubik and Botman (1976) have also made several studies of the accuracy of digital terrain models using different interpolation techniques. Their accuracy models are related to point density, interpolation techniques and terrain characteristics, which were assumed to be known and described in terms of **(auto)covariance.** Again, as Ackermann (1979) pointed out, "up to now, the theoretical results have not yet been compared with empirical results". Therefore, "a study would be required to see how well covariance functions can describe real terrain and which function should be assigned to different types of terrain". Yet again, such a study as suggested by Ackermann (1979) has not yet been carried out up to the present time - at least to the present author's knowledge.

In a manner similar to their work of using covariance, Kubik and his collaborators (Frederiksen et al, 1983; 1986) also tried later to use the **variogram** which is a basic concept in geostatistics (see Chapter 3). They connected this variable to the covariance which was used by Kubik and Botman (1976) to produce yet another model for DTM accuracy prediction. As one can imagine, again a study would be required to see how well the variogram functions can describe real terrain and which function should be assigned to different types of terrain, although in the papers published by Frederiksen et al (1983, 1986), some examples were given, based on very limited tests.

Frederiksen (1980) and his colleagues (Frederiksen et al., 1978) have also designed a mathematical model on the basis of the **summation of Fourier spectra** of the terrain surface in their high-frequency parts, i.e. those higher than 1/(2.Dx), where Dx is the sampling interval. Here, again, the Fourier spectra of a terrain surface is assumed to be known or estimated using measured profiles comprising data points of a high density. Then the accuracy of the DTM is estimated by summing the known spectra from the frequency of 1/(2.Dx) to the infinite, which is supposed to be the part of information loss due to the sampling with an interval of Dx. However, one important point which has not been taken into account in their model is the fact that the magnitude of the spectra will also be reduced if the spectra are computed from data points with a larger interval between these points. As a result, their model may well produce too optimistic a prediction.

12.3.2 Experimental evaluations of existing accuracy models

Several mathematical models of DTM accuracy have been outlined briefly in the previous section (12.3.1), and it has also been pointed out that a study of these various models is required to see how well they can predict DTM accuracy. Indeed, this section is an attempt to carry out such a study. Actually, the evaluation has been limited to the models based on variogram analysis, (auto)covariance analysis and the summation of Fourier spectra over their high-frequency parts.

The data sets used for this evaluation are again those generated for the ISPRS DTM test which have been described in Chapter 9.

(1). Evaluation of the models based on variogram function

The mathematical expression of this model can be expressed as follows:

$$VAR(Int) = A \left(\frac{Dx}{L} \right)^{b} \left(\frac{1}{3} + \frac{4}{(b+1)\,(b+2)} \right) \qquad (12.1)$$

Where, A and b are the parameters of the variogram functions given in Equ.(3.10); L is the sampling interval along the profiles which were used to compute the parameters A and b; Dx is the sampling interval of the DTM source data; and VAR(Int) denotes the DTM accuracy without taking the errors in the DTM source data into consideration.

Table 12.1 Regression results for the coefficients for the model based on variogram analysis

Test Area	L	2 A	b	r
Uppland	40m	5.32	1.196	0.999
Sohnstetten	20m	17.15	1.727	0.998
Spitze	10m	1.365	1.537	0.9998

N.B. r is the correlation coefficient for the regression.

Table 12.2a Comparison of theoretical values produced by Equ.(12.1) with experimental test results obtained from gridded data sets

Test Area	Dx	Dx / L	Predicted	Test Result	Differ.
Uppland	40m	1.0	0.63m	0.76 m	-0.14m
	56.56m	1.414	1.18m	0.93 m	0.25m
	80m	2.0	1.38m	1.18 m	0.20m
Sohnstetten	20m	1.0	0.52m	0.56 m	-0.04m
	28.28m	1.414	0.98m	0.87 m	0.11m
	40m	2.0	1.38m	1.45 m	-0.07m
	56.56m	2.828	1.77m	2.40 m	-0.63m
Spitze	10m	1.0	0.08m	0.21 m	-0.13m
	14.14m	1.141	0.37m	0.28 m	0.09m
	20m	2.0	0.48m	0.35 m	-0.13m

The semi-variograms computed from different data sets are shown in Fig.12.1(b), (d) and (f). For each area, the variogram produced from the data set with the smallest sampling interval has been used to

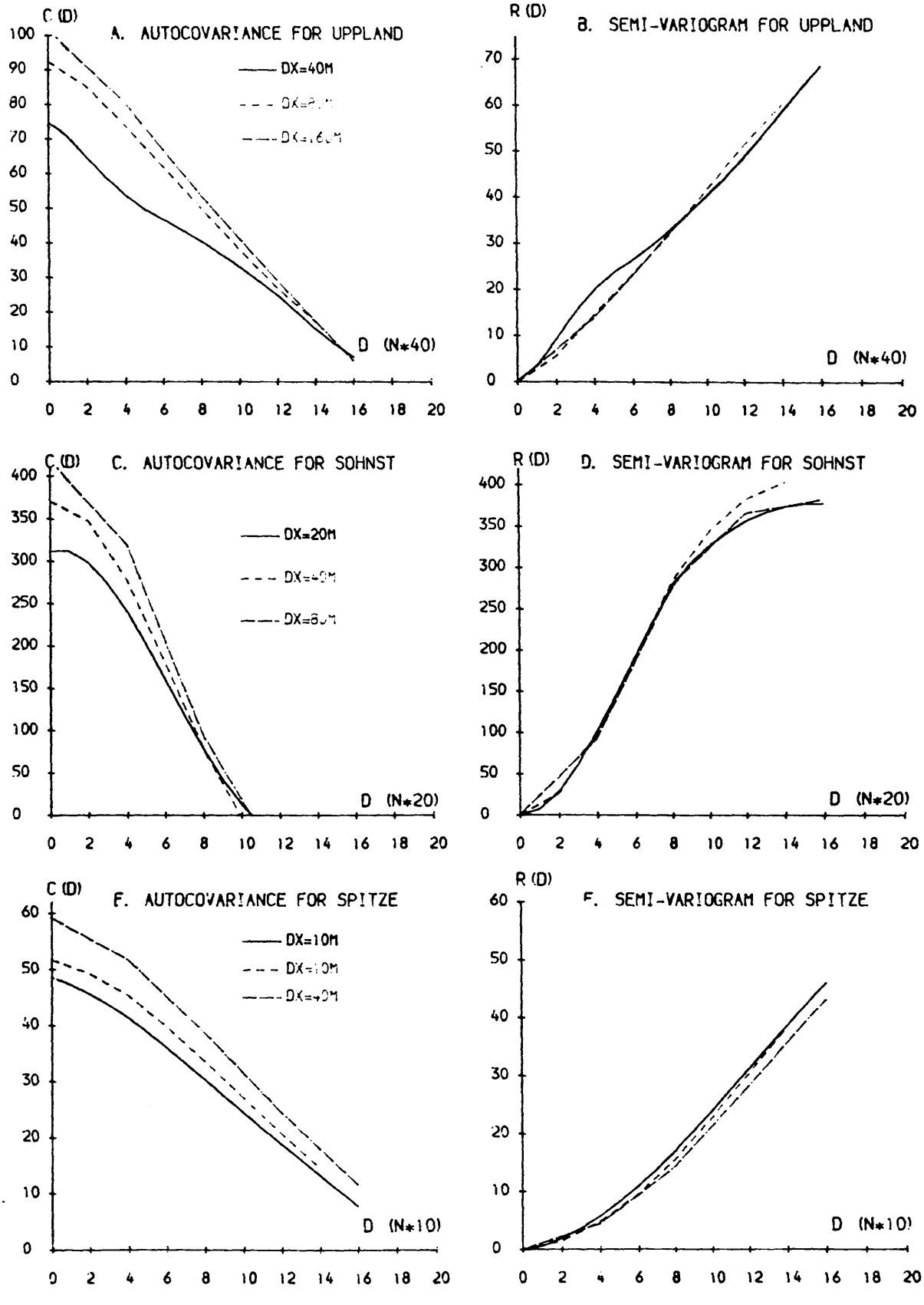


FIG.12.1 COVARIANCE AND SEMI-VARIOGRAM WITH SAMPLING DISTANCE

compute the coefficients for these models using a regression technique. Actually, for the Sohnstetten area, only the first 8 points were used for the regression. The results are given in Table 12.1. Also some examples of the comparison of the accuracy results predicted by this model with the experimental results which were described in Chapter 9 are shown in Table 12.2a.

It can be seen that the predicted results are mostly in reasonable agreement with the actual results obtained from the experimental test data with the vast majority of the differences being below 0.13m. The poorer results are all associated with larger sampling intervals. For example, for the Uppland area, the values 0.25m and 0.20m were obtained for the 56.56m and 80m sampling intervals respectively, while for the Sohnstetten area, the largest difference is up to 0.63m for the case of 56.56m interval.

It needs to be pointed out here that the values listed under the heading "test result" are the average values obtained from the gridded data sets. However, when the DTM accuracy figures obtained from the composite data sets were used, the differences were found to be much greater. A comparison of these results is given in Table 12.2b. From this it can be seen that some results are unacceptable in the case of the DTM derived from composite data sets.

Table 12.2b Comparison of theoretical values produced by Equ.(12.1) with experimental test results obtained from composite data sets

Test Area	Dx	Dx / L	Predicted	Test Result	Differ.
Uppland	40m	1.0	0.63m	0.66 m	-0.03m
	56.56m	1.414	1.18m	0.70 m	0.48m
	80m	2.0	1.38m	0.80 m	0.58m
Sohnstetten	20m	1.0	0.52m	0.43 m	0.09m
	28.28m	1.414	0.98m	0.56 m	0.42m
	40m	2.0	1.38m	0.78 m	0.60m
	56.56m	2.828	1.77m	1.08 m	0.69m
Spitze	10m	1.0	0.08m	0.16 m	-0.08m
	14.14m	1.141	0.37m	0.17 m	0.20m
	20m	2.0	0.48m	0.18 m	0.30m

Analyzing the results, it will be seen that the difference values were very small in the case of the smallest sampling interval used for each of the three test areas - -0.03m in the case of Uppland; 0.09m in the case of Sohnstetten; and -0.08m in the case of Spitze. However, in the case of the Uppland area, as soon as larger sampling intervals were used, the difference values increased to 0.48m (for 56.56m) and 0.58m (for 80m). A similar experience was noted with the Sohnstetten area, where the values increased to 0.42m (for 28.28m); 0.60m (for 40m) and 0.69m (for 56.56m) respectively. In the case of the Spitze area, the increase in the difference values was a litte less marked with 0.20m (for 14.14m) and 0.30m (for 20m), but these values are still sharp increases in the value (-0.08m) for the minimum sampling interval of 10m.

It is very interesting to note that the predictions produced by this variogram model are so close to the actual results obtained from the gridded data sets, whereas they are not at all in good agreement with the results obtained from the composite data sets. This might be due to the fact that the values of the variogram used in this model were computed from gridded data sets only and not from the composite data sets since it is complicated and difficult to compute variograms from non-gridded data sets. As one can imagine, if the variogram values were computed from grids with very small grid intervals, then the prediction produced by this model might be closer to the results obtained from composite data sets.

Another interesting point is that, in the case of gridded data, the difference between the experimental results and the prediction produced by this model is rather consistent with the increase in sampling interval. This is due to the fact that the semi-variograms (see Fig.12.1) computed from the gridded data with different grid intervals were very similar to one another.

(2). Evaluation of the model based on covariance function

The same ISPRS test data has also been used to evaluate the models by means of covariance analysis. Here only one model has been selected - that for linear interpolation using the exponential covariance model. The expression is follows:

$$SD(int)_{max} = 1.1\ SD(T)\ (Dx/c)^{1/2} \qquad (12.2)$$

Where, $SD(int)_{max}$ is the maximum standard error due to topographic generalisation; SD(T) is the standard error of the terrain height, i.e. the square root of the terrain variance; Dx is the grid interval (or sampling interval of the source data; and c is the parameter used in the variance model. Of course, $SD(int)_{mean}$ should be used in practice.

However, the difference is very small as shown by Kubik and Botman (1976). Therefore, Equ.(12.2) has been used in this particular case.

Considering the effect of errors in the source data, the following expression was suggested by Kubik and Botman (1976):

$$SD(DTM) = (SD^2(int) + 0.36\ VAR(raw))^{1/2} \qquad (12.3)$$

Where, the VAR(raw) is the accuracy of the raw (source) data and SD(DTM) is the predicted accuracy of the DTM.

The (auto)covariances computed from the test data are shown graphically in Figs.12.1(a), (c) and (d). The following formula was used to carry out the regression to estimate the parameters for (12.2):

$$Y = A\ Exp(B.X) \qquad (12.4)$$

The regression results using Equ.(12.4) are shown in Table 12.3.

Table 12.3 Regression results for the coefficients for the model based on covariance analysis

Test Area	L	A	B	r
Uppland	40m	76.178	-0.002338	-0.988
Sohnstetten	20m	327.540	-0.005285	-0.944
Spitze	10m	51.684	-0.007848	-0.978

N.B. r is the correlation coefficient for the regression.

A comparison of the results predicted by this mathematical model with the experimental results is shown in Table 12.4.

From these two tables, it can be seen that the predictions are quite poor. The results might be improved by using the Gaussian variance model instead of the exponential model (Kubik and Botman, 1976). However, as can be seen from Fig.12.1, the covariance values computed from the gridded data with different intervals were so different that it is difficult for this model to produce a very consistent prediction no matter which model is used to express the covariance value. As one can see from Fig.12.1, clearly there is no indication to show that the covariance values can be expressed better by the Gaussian function. Therefore, such an investigation is omitted here since an exhaustive evaluation of existing accuracy models is not the main purpose of this project.

Table 12.4a Comparison of accuracy predicted by Equ.(12.3) with test results obtained from gridded data sets

Test Area	Dx	SD(int)	Predicted	Test result	Differ.
Uppland	40m	0.30m	0.50m	0.76 m	-0.26m
	56.56m	0.35m	0.53m	0.93 m	-0.40m
	80m	0.42m	0.58m	1.18 m	-0.60m
Sohnstetten	20m	1.02m	1.02m	0.56 m	0.46m
	28.28m	1.22m	1.22m	0.87 m	0.35m
	40m	1.45m	1.45m	1.45 m	0.00m
	56.56m	1.72m	1.72m	2.40 m	-0.68m
Spitze	10m	0.49m	0.49m	0.21 m	0.28m
	14.14m	0.59m	0.59m	0.28 m	0.31m
	20m	0.70m	0.70m	0.35 m	0.35m

Table 12.4b Comparison of accuracy predicted by Equ.(12.3) with test results obtained from composite data sets

Test Area	Dx	SD(int)	Predicted	Test result	Differ.
Uppland	40m	0.30m	0.50m	0.66 m	-0.16m
	56.56m	0.35m	0.53m	0.70 m	-0.17m
	80m	0.42m	0.58m	0.80 m	-0.22m
Sohnstetten	20m	1.02m	1.02m	0.43 m	0.59m
	28.28m	1.22m	1.22m	0.56 m	0.66m
	40m	1.45m	1.45m	0.78 m	0.67m
	56.56m	1.72m	1.72m	1.08 m	0.64m
Spitze	10m	0.49m	0.49m	0.16 m	0.33m
	14.14m	0.59m	0.59m	0.17 m	0.42m
	20m	0.70m	0.70m	0.18 m	0.62m

(3). Evaluation of the model based on high-frequency spectral analysis

As mentioned previously, to make use of the mathematical models based on Fourier spectral analysis, digital profile data with a high density must be used. In the case of the gridded data from the ISPRS test, it was considered that the grid data points were not sufficiently dense for such a purpose. However, such estimation work using the model based on the summation of Fourier spectra over the high-frequency part of a data set has been carried out by Prof. Ackermann and Dr. Sigle at the University of Stuttgart. The results predicted by this model for the ISPRS data sets measured at the University of Stuttgart, together with the original data sets, were made available to the author through the courtesy of Prof. Ackermann and Dr. Sigle.

The data sets measured at the University of Stuttgart included those for three of the ISPRS areas - Sohnstetten, Spitze and Drivdalen. A brief description of the terrain features existing in the first two areas has already been given in Chapter 9 and therefore will not be repeated here. The Drivdalen area has a very steep and rough terrain surface with a height variation of 780m and an average slope of about 40^{o}. The height accuracies (RMSE values) of the source data sets for these test areas were estimated as $\pm$0.17m for the Sohnstetten area; $\pm$0.10m for the Spitze area; and $\pm$0.35m for the Drivdalen area. These data sets have also been processed by the author using the same procedures as described in Chapter 9 and the results are given in Table 12.5.

Table 12.5a Accuracy results for the gridded data sets from Stuttgart

Test Area	Data Set	RMSE	SD	MEAN	+Emax	-Emax
D. Drivdalen	20m Grid	1.61	1.57	-0.36	9.89	-14.78
E. Sohnstetten	15m Grid	0.49	0.46	-0.15	2.70	-3.51
F. Spitze	15m Grid	0.31	0.31	0.05	4.01	-1.80

Table 12.5b Accuracy results for the composite data sets from Stuttgart

Test Area	Data Set	RMSE	SD	MEAN	+Emax	-Emax
D. Drivdalen	20m Grid +P	1.50	1.47	-0.34	9.89	-9.92
E. Sohnstetten	15m Grid +P	0.39	0.35	-0.15	1.44	-2.20
F. Spitze	15m Grid +P	0.20	0.20	0.03	4.65	-1.08

A comparison of the predicted values with the results obtained from the experimental tests has been given in Table 12.6. In general, it can be found that the results are not at all bad. Indeed, in the case of the composite data, the results can be considered as good. However, it must be pointed out here that these results were obtained from the data sets with small sampling intervals.

Table 12.6 Comparion of test results with the value predicted by the model based on the summation of high-frequency Fourier spectra

Test Area	Grid	Predicted	Gridded data		Composite data	
			Tested	Differ.	Tested	Differ.
Sohnstetten	15m	0.26m	0.46m	-0.20m	0.35m	-0.09m
Spitze	15m	0.10m	0.31m	-0.21m	0.20m	-0.11m
Drivdalen	20m	1.25m	1.57m	-0.32m	0.47m	-0.22m

It will be noted that the results predicted by this model are quite close to the accuracy figures obtained from composite data sets. This may be due to the fact that the spectra in this case were computed from high-density profile data. In such a case, it may well be the case that the information about feature-specific points and lines was more or less included in the profile data. Thus the results produced by this model are very close to those obtained from the composite data sets. It is also interesting to note that, in the case of these limited results, this model, as expected, always produces too optimistic a prediction.

12.3.3 Discussion

In this section, three existing mathematical models of DTM accuracy have been evaluated experimentally. From the comparisons of the values predicted by these models with the test results, it can be found that these models behave very differently. In general, the predictions made by the models based on the use of the covariance function are quite poor. On the other hand, the results predicted by the model based on the variogram function seem to be quite similar to those obtained from gridded data sets but not with those resulting from the composite data sets. By contrast, the results predicted by the model based on spectral analysis are closer to those obtained from the composite data sets.

It must be pointed out here that the parameters of the models based on both variogram analysis and covariance analysis were estimated from the whole set of data points. In practice, it is impossible to do this with confidence since the DTM accuracy for a given sampling interval needs to be predicted before the actual measurement of the data points can be

carried out. Therefore, it may be not easy to obtain very reliable estimates for the parameters in these models. Thus, these models may be not too practical even though some of them may produce reasonable results in some cases.

As noted by Ackermann (1979), in practice, to make use of the models based on covariance (or variogram) analysis, an intensive study of the covariance (or variogram) for different types of terrain is necessary in order to make use of these models for DTM accuracy prediction. However, the covariance (or variogram) values for different areas even of the same geomorphological type may be different. Even for the same area, the covariance values derived for different directions could be quite different. This means that it may still be very difficult for these models to produce a very reliable prediction. For those models based on a Fourier spectrum analysis, the situation is very similar to those based on covariance or variogram functions.

According to these experimental results and the theoretical discussion conducted above, it can be concluded that these models cannot satisfy all the seven standards for evaluating mathematical models which have been discussed in Chapter 2 .

Considering the demerits of these numerical descriptors such as Fourier spectrum, variogram, covariance, etc (see Chapter 3), and the relative weakness of those mathematical models of DTM accuracy based on these descriptors, it can be concluded that, at least on the present rather limited evidence, these models may not be among the best models that can be used for DTM accuracy prediction. Therefore, attention should be paid to the development of alternative mathematical models which should be more appropriate both in theory and in practice. The parameters of such a model must have a physical realisation. Such models will be discussed in the next sections in this chapter.

12.4 Background information about the new DTM accuracy model

In the previous sections (12.2 and 12.3), the basic approaches for DTM accuracy assessment have been briefly discussed; the existing accuracy models have been outlined; and some of the accuracy models have also been evaluated by experimental tests. From this section onwards, the discussion will be concentrated on the development of alternative mathematical models for the accuracy of the DTMs derived from photogrammetrically measured data. Before starting the derivation of formulae, the basic information about the procedures for the model development will be described in this section.

12.4.1 Basic parameters in the new accuracy model

Before the procedure used for accuracy model development can be described, first of all, the parameters which will be used in this accuracy model need to be specified here.

As has been identified previously (in Chapters 1 and 9), the accuracy of a digital terrain model surface is affected by **six main factors,** i.e. the characteristics of the terrain surface; the three attributes of the DTM source data; the method used for DTM surface modelling, and the characteristics of the DTM surface itself.

The characteristics of the **terrain surface** in this model will be described mainly by the mean (or representative) value of the slope angles in the areas to be modelled since the use of slope has many advantages over other numerical descriptors (see Chapter 3). The letter "A" will be used to denote the mean slope value.

As has also been discussed in Chapter 3, there is a high correlation between the slope value and the variations in the elevation values. Therefore, relief information may be used as a rough guide. Based on the experience accumulated by the surveying and mapping communities in many countries, a general relationship between the slope values and the relief information can be summarized in Table 12.7.

Table 12.7 General relationship between relief and slope angles

Type of Terrain	Local Relief	Mean Slope (^{o})
Flat regions	< 20 m	< 2
Hilly areas	< 150 m	2 - 6
Low mountains	> 150 m	6 - 25
High mountains	> 500 m	> 25

The DTM **source data** which will be considered in this study is the photogrammetrically measured data. As has been discussed in Chapters 4 and 5, three attributes i.e. accuracy, distribution and density, are associated with such a data set.

The **accuracy** of such photogrammetrically measured data has been discussed in Chapter 5. It has also been summarized there that $\pm$0.067 to 0.20 per mil of H is the range of RMSE values which are commonly accepted for the accuracy of data points measured in a static mode and $\pm$0.3 per mil of H might be a reasonable value for the data points

measured in a dynamic mode - although, in fact, the tests of the Sudan model gave a much better value. In this study, the accuracy of the DTM source data is expressed in terms of either variance, denoted as VAR(PMD), or standard deviation, SD(PMD).

The **distribution** of the data is specified by the **data pattern.** The patterns which will be considered in this study are limited to square gridded data and a composite of the square grid and the data about the feature-specific points and feature-specific lines, since these are the most widely used ones. Actually, the orientation of the data pattern is also associated with the data distribution and this may also have effects on the accuracy of the final DTM. However, this factor has not been taken into consideration here due to the complexity of the matter itself.

The **density** of the DTM source data is specified by its grid interval since the data points are homogeneous over all of the area. Dx is used to denote the grid interval, and it is also referred to as the sampling interval.

The **method used for surface modelling** which will be considered in this study is limited to one which constructs a DTM surface directly from the acquired source data. This means that no random-to-grid interpolation procedure has been applied to the source data, thus there is no loss in accuracy in the source data due to pre-processing. By considering only this method, the complex matter of the loss in the fidelity of representation of the terrain topography due to the random-to-interpolation process can be avoided, since inevitably such a process will have some generalisation effects on the spatial variations present in the source data.

The type of **DTM surface** which will be considered in this study is limited to the continuous surface comprising a series of linked linear facets which are either triangular facets or bilinear surfaces in the case of gridded data or a hybrid of a series of contiguous bilinear surfaces and triangular facets in the case of composite data since, as has been pointed out before, surfaces formed from such data are the least misleading in most cases. Another reason why only this type of continuous surface is being considered is for the sake of the resulting simplicity in analysis.

12.4.2 The line of thought and procedure for the model development

After the description of this background information about the basic parameters used in this model, the line of thought for the model development can be outlined here, then the procedure which has actually been used for the model development can be described here.

It is obvious that the errors at the DTM points are the results of the **deviations (discrepancies) of the DTM surface** from the real terrain surface. These deviations result in a loss in the fidelity of the terrain topography represented by the DTM. This can be considered to come from **two basic aspects:-**

i). the errors present in and propagated from the source data points; and

ii). the generalisation effect resulting from the sampling process during acquisition and the subsequent reconstruction carried out by the modelling program.

The degree of this **generalisation** is a function of the sampling interval of the source data, the characteristics of the DTM surface and the method of surface modelling. Also the distribution of the source data, the characteristics of the terrain surface, and even the errors at the source data points contribute to this effect. The errors in the source data points will of course be propagated to the DTM points.

Therefore, the **procedure** used in this study will be, first of all, to investigate the possible magnitude of the errors which may come from the individual items of the two basic aspects mentioned above, and then to integrate the errors from these two different sources into a single entity.

12.5 Propagation of the errors from the source data

The discussion of the DTM errors will start with the propagation of errors from the DTM source data. Before this discussion can be carried out, the interpolation of the data points occuring on the DTM surface needs first of all to be considered. Of course, these interpolation procedures are not at all new. However, a description of the interpolation will facilitate the theoretical analysis of the error propagation. Indeed, it will be found that such a description is necessary for this purpose.

12.5.1 Interpolation of data points on a continuous DTM surface

Since the type of DTM surface which will be considered in this study is confined to the continuous surface comprising a series of contiguous linear facets, therefore, the interpolation of the data points carried out on such a DTM surface will be a linear interpolation.

(1). Linear interpolation between two points

Linear interpolation between two points is the simplest possible procedure. The principle is illustrated in Fig.12.2. In the figure, H_I is the height of the interpolated point I; H_A and H_B are the heights at points A and B; Dx denotes the sampling interval (spacing); and X is the distance between the point to be interpolated and point A (in plan). The mathematical function for the linear interpolation between A and B is as follows:

$$H_I = (H_A\ (Dx-X) + H_B\ X) / Dx \qquad (12.5)$$

(2). Interpolation in a bilinear surface

The interpolation in a bilinear surface which may be constructed from a series of contiguous square grids is illustrated in Fig.12.3. In this figure, A, B, C and D are four nodes of the grid, from which the bilinear surface is constructed. Point I is the point to be interpolated. In order to determine the height of point I, the heights of E and F (or G and H) need first of all to be interpolated using Equ.(12.5). Therefore, the interpolation of a point, I in this case, in a square grid is as follows:

$$H_I = (H_H\ (Dx-X) + H_G\ X) / Dx \qquad (12.6)$$

Where,

$$H_H = (H_C\ (Dx-Y) + H_D\ Y) / Dx \qquad (12.7)$$
$$H_G = (H_B\ (Dx-Y) + H_A\ Y) / Dx$$

(3). Interpolation in a right-angled isosceles triangular facet

As discussed in Chapter 7, the square grid data can also be split into a triangular network, which consists of a set of isosceles right-angled triangles. From this network, again a continuous surface comprising a series of contiguous linear facets can be constructed.

There are many ways to interpolate a point in an isosceles right-angled triangular facet. Fig.12.4 shows some of these possibilities. In this figure, A, B, and C are the three nodal points and I is the point to be interpolated. Fig.12.4(b) is the plan view of the perspective diagram (a). From Fig.12.4(b), it can be found that I can be determined by A and G, or J and C, or B and E, or H and D, or K and F. In this particular case, purely for the convenience of the later discussion of the matter of the generalisation of terrain topography due to the sampling pattern and the later reconstruction, the interpolation using points K and F, or D and H is discussed, since basically this is similar to the interpolation procedure carried out in a square grid. By employing the procedure discussed in the preceding section, the

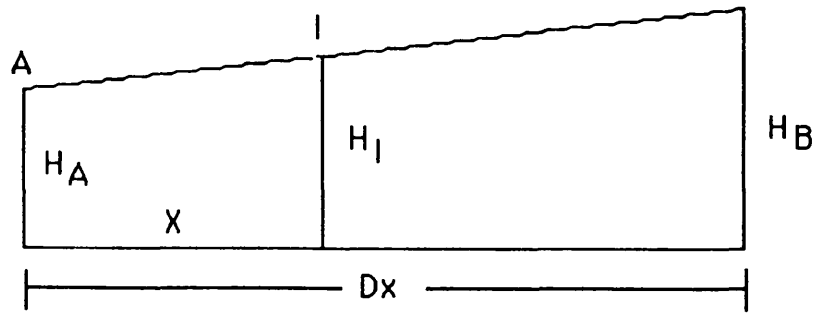


Fig.12.2 Linear interpolation between two points

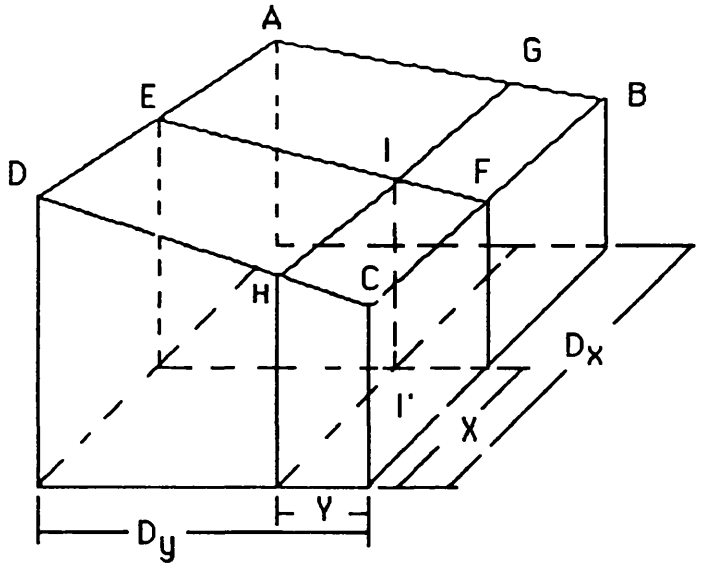


Fig.12.3 Linear interpolation in a square grid

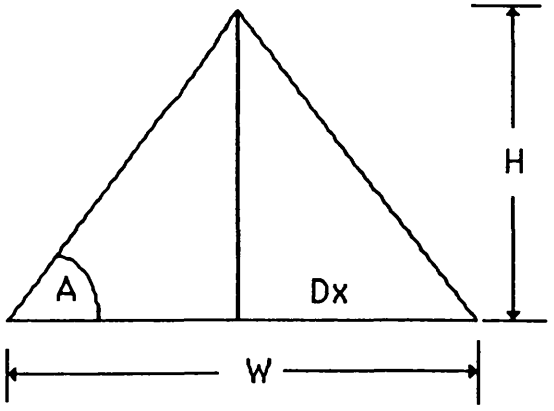
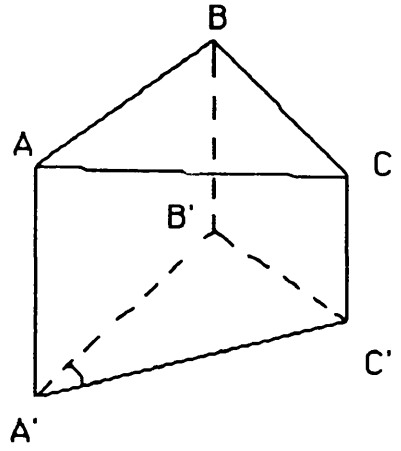
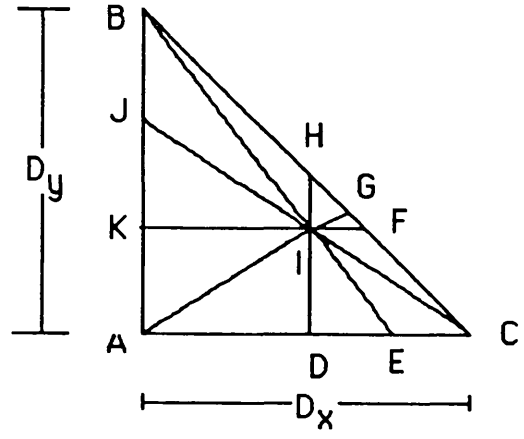


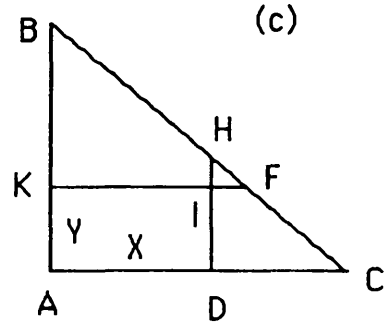
Fig.12.6 Rough estimation of W



(a)



(b)



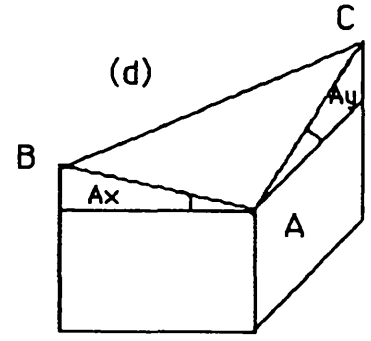


Fig. 12.4 Linear interpolation in a isoceles triangle

following equations can be obtained (Fig.12.4 (b) & (c)):

$$H_D = (H_A\ (Dx-X) + H_C\ X) / Dx \qquad (12.8)$$
$$H_H = (H_B\ (Dx-X) + H_C\ X) / Dx$$

Finally, the height at interpolation point, I in this case, is:

$$H_I = (H_D\ (Dx-Y) + H_H\ Y) / Dx \qquad (12.9)$$

The slope angle of BC, hy, is not equal to that of either AC or AB. It is illustrated in Fig.12.4.(d). It can be found that:-

$$\tan(hy) = (\tan(Ax) + \tan(Ay)) / 2^{1/2} \qquad (12.10)$$

Where, Ax denotes the slope angle in one direction, say the profile direction, in which case, Ay denotes the slope angle in the direction perpendicular to the profiles. The maximum value for tan(hy) is $2^{1/2}$ tan(Ax) when tan(Ax) is equal to tan(Ay) and the minimum value is tan(Ax)/$2^{1/2}$ when tan(Ay) is equal to zero. The average value of tan(hy), considering different cases, is therefore as follows:

$$\begin{aligned} \tan(hy)av &= (\ 2^{1/2}\tan(A) + \tan(A)/\ 2^{1/2}\ )/2 \\ &= 3/4 \times 2^{1/2}\ \tan(A) \\ &= 1.06\ \tan(A) \end{aligned} \qquad (12.11)$$

Where A denotes the average slope angle in the area.

12.5.2 Propagation of errors from the source data to DTM data

Turning next to the propagation of the errors in the PMD (photogrammetrically measured data) to the data points interpolated on the DTM surface, the accuracy of the PMD can be expressed by a variance, VAR(PMD) or standard deviation, SD(PMD). By applying the error propagation laws to Equ.(12.5), the variance of the errors at the point interpolated in the case of a linear profile, VAR(P) can be obtained as follows:

$$Dx^2\ VAR(P) = (Dx-X)^2\ VAR(PMD) + X^2\ VAR(PMD) \qquad (12.12)$$

This is the expression for a point located at a distance X from a grid node and (Dx-X) from the other grid node. In practice, it is very difficult to deal with individual points. Furthermore, the result which is of interest to both DTM producers and DTM users is the overall average value. Therefore, it is desirable that the average value of VAR(P) be used as the loss of accuracy arising from the errors propagated from the PMD. This average value can be computed by summarizing all the possible VAR(P) values over a grid and dividing

them by its sampling interval (Dx). Such an operation can be expressed as follows:

$$Dx^2\ VAR(P) = \left(\int_0^{Dx} ((Dx-X)^2\ VAR(PMD) + X^2\ VAR(PMD))dx \right)/Dx \qquad (12.13)$$

Therefore, $VAR(P) = 2/3\ VAR(PMD) = 0.667\ VAR(PMD)$

or $SD(P) = (2/3)^{1/2} SD(PMD) = 0.82\ SD(PMD)$ (12.14)

This represents the propagation of errors from the source data to the DTM points in one direction only. In a similar way, the propagation of errors in the other direction can also be evaluated. Considering the case of points H and G in Fig.12.3, the effects of errors in the source data on the accuracy of points H and G can be estimated using Equ.(12.14). When point I (Fig.12.3) is to be interpolated, the errors at points H and G will affect the accuracy of point I. Therefore, Equ.(12.13) can also be used to estimate the accuracy loss due to the errors propagated from points H and G. However, the accuracies of points H and G will be higher than that of any of the grid nodes (i.e. either A, B, C or D). In this case, the generalisation of terrain topography along the profile lines of DC and AB will bring a loss of accuracy to points H and G, respectively. It is clear that the error propagation is a nested process. Therefore, before any further progress can be made, the effect of generalisation of the terrain topography on point accuracy should be studied.

12.6 Accuracy loss of DTM data points due to generalisation effect

The effect of generalisation of the terrain topography on the accuracy of the data points interpolated on a DTM surface is usually more serious than the errors in the source data, especially when a large sampling interval is in use. Therefore, it is extremely important to understand the effects of such a generalisation. This section is an attempt to carry out such an investigation.

12.6.1 Introduction

From DTM literature, it can be found that the investigation into the generalisation effect has been carried out in different ways such as using Fourier spectra, covariance, variogram, etc. However, as has been pointed out previously (Section 12.3), the applicability of those works is somewhat limited. Therefore, an attempt is made here to relate the generalisation effect to the slope angle of the terrain surface. The **procedure** used in this section is as follows:

i). First of all, the **two extreme values** (the positive maximum and negative maximum errors) of the errors resulting from the loss in the fidelity of the terrain surface caused by the generalisation effect due to sampling and reconstruction are analysed;

ii). Then, the **relationship** between the two extreme values and the standard deviation of the distribution of DTM errors due to the generalisation effect will be investigated;

iii). Finally, the two extreme values are used for the estimation of the value of the **standard deviation** which represents the accuracy loss due to the generalisation effect, for different types of data.

12.6.2 Analysis of extreme errors

In order to analyse the possible value for the **extreme errors**, the possible shapes of terrain **profiles** need to be considered. Fig.12.5 shows some of the possible shapes of such profiles and the ways in which extreme errors may arise due to the linear construction of DTM surface from the gridded data points.

Fig.12.5(a) shows the maximum possible error at point C due to a **fault** or other geological structure giving rise to the steep change in slope. Therefore, if the information giving a full description of this structure or discontinuity has not been collected, a huge error may result. The value of such an error, E_b, varies with the characteristics of the feature itself. It can only be measured and is not readily predictable. In this specific diagram, the value is approximately as the following:

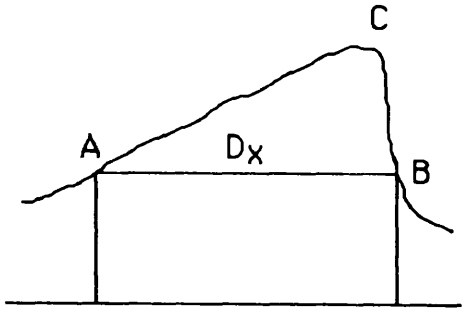
$$E_b = CB = Dx \tan A \qquad (12.15)$$

Where A is the slope angle at point A (i.e. angle CAB) and Dx is the sampling interval (spacing). This value might be used as a representative value when E_b is relatively small.

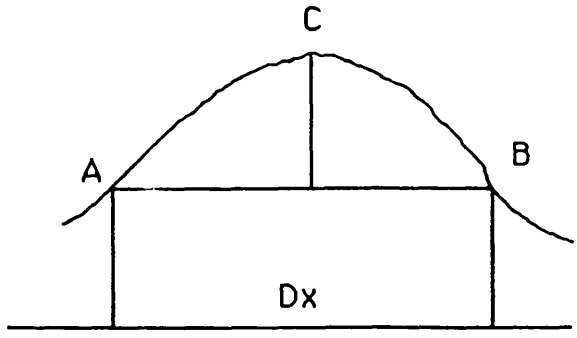
Fig.12.5(b) shows the possible maximum error at a point C where **regular grid sampling** has been carried out without selecting local minima and maxima, i.e. feature-specific points and the points along feature-specific lines have not been measured. The possible value of this maximum error when point C lies at the top of the hill represented by the diagram is as follows:

$$E_r = (1/2)\ Dx \tan A \qquad (12.16)$$

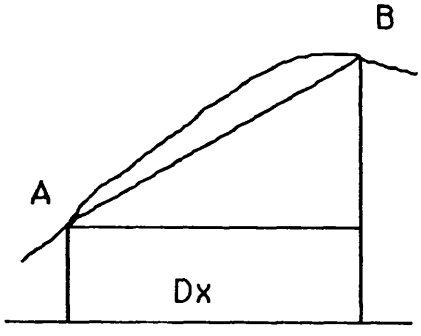
Figs.12.5(c) and (e) also show some other possible profiles formed by



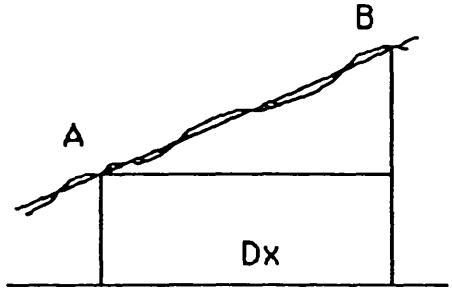
(a) with a discontinuity



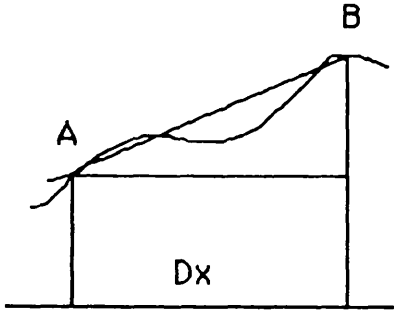
(b) maximum error using a regular grid



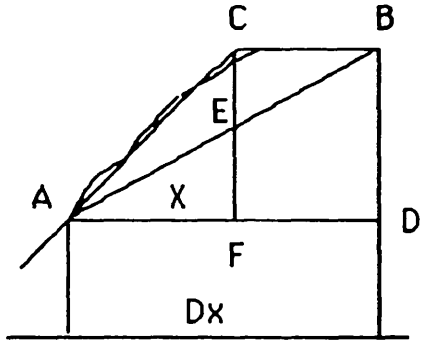
(c) a convex slope



(d) a regular slope



(e) a concave slope



(f) Error due to linear interpolation

Fig.12.5 Different type of slopes

convex and concave slopes respectively and the possible errors which would result when a linear surface is constructed from them. Fig.12.5(d) is yet another case, where the ground has a regular slope and the profile is almost a straight line.

Fig.12.5(f) shows the possible maximum error which may occur for the data obtained from **composite sampling.** Here it has been assumed that the data points are gridded data plus local maxima and minima. That is to say, the points along ridge lines and ravine lines are included, but the points defining convex and concave slopes are still not all included. This assumption seems to be not too unrealistic since it is not practical to include all the convex and concave points except in the case when pure selective sampling has been carried out on a stereo-model to simulate the data acquisition by ground surveying. On the other hand, the diagram of Fig.12.5(f) itself may seem very unrealistic since the point C in this case will be selected in practice. Indeed, it must be admitted that it is quite exaggerated. However, such an exaggeration is necessary since it is the extreme case that is under investigation in this study.

After this assumption, the extreme errors can be analysed using Fig.12.5. In this diagram, the line AB is the linearly constructed profile. It can be seen clearly that the largest error is at point C and the value of CE will be analysed below. In this diagram, angle CAD is the slope angle, A; so CF = BD = X tanA; and EF=BD.X/Dx; so

$$CF = X.\tan A - (X.\tan A).X/Dx = X.\tan A - X^2 \tan A / Dx \qquad (12.17)$$

By letting the first derivative of CF be equal to zero, the location of X where the value of CF reaches its maximum can be determined. So,

$$d(CF)/dX = \tan A - 2.X.\tan A / Dx = 0 \qquad (12.18)$$

From Equ.(12.18), it can be found that X = Dx/2. This means that the position where CF reaches its maximum value is in the middle of a grid. By letting X=Dx/2, the maximum value of CF, donoted as E_c, can be found from Equ.(12.17) as follows:

$$E_c = Dx \tan(A) / 4 \qquad (12.19)$$

Therefore, it can also be noticed that the value of possible extreme errors in the case of only regularly gridded data is 2 times that in the case of composite data. It needs to be pointed out that the extreme values estimated in this section are all positive maximum values. The estimation of the corresponding negative maximum values can be carried out in a similar manner but obviously will result in absolute values of the estimated negative maximum errors being equal to those of the

corresponding positive maximum errors. Therefore, such an estimation will not be duplicated here in this study.

12.6.3 Relationship between extreme errors and the standard deviation of the distribution

In DTM practice, the accuracy figures are normally expressed in terms of RMSE (root mean square error) or SD (standard deviation). Therefore, these extreme values of DTM errors caused by the generalisation of the terrain topography represented by the DTM surface need to be converted to the SD values. This section is such an attempt.

If the **distribution** of the errors due to the generalisation effect is known, then it will be an easy task to convert the values of the extreme errors into the value of the standard deviation. However, the problem in the context of DTM accuracy assessment is that the distribution of DTM errors is still not known. Torlegard et al (1986) found that "they (the unfiltered DTM errors) are not normally distributed with a zero mean. After filtering out the blunders and levelling the DEMs, quite a number of them seem to show a normal distribution". This statement implies that DTM errors may approximate to a normal distribution, but are not necessarily normal. This means that the laws for normal distribution may be not applicable in the case of a DTM. Therefore, some alternative approaches need to be considered.

Nevertheless, for any distribution, according to **Chebyshev's theorem**, most of the observations of a random variable (DTM errors in this case) of will fall within the range of 4 times the SD (standard deviation) distant from u (mean). Chebyshev's theorem states that the probability is at least as large as $1-1/k^2$ that an observation of a random variable, say X, will be within the range from u-k.SD(X) to u+k.SD(X), i.e.

$$P(|X-u| > k.SD(X)) < 1/k^2 \qquad (12.20)$$

or

$$P(|X-u| \leq k.SD(X)) \geq 1-1/k^2 \qquad (12.21)$$

Where, k is any constant larger than 1. If the distribution is approximately normal, the probability of a point lying within this range is much larger than the value P in Equ.(12.21).

More importantly, from the **experimental test results** which were given in Chapter 9, it was found that the occurrence frequency of DTM residuals (errors) larger than 4xSD is usually much smaller than 1.0% in most cases. Therefore 4xSD in this case, like 3xSD in the normal distribution, can be used as the limit of the maximum random errors.

From the discussion above, it can be concluded that from both the practical and theoretical points of view, the value of **4xSD** can be used as the criterion for the maximum random error. Therefore, conversely, if the maximum random error is known, then the standard deviation of the distribution of the errors can also be estimated. Thus, the possible maximum random errors which have already been estimated in the previous section (12.6.2) can be converted into a standard deviation. This will be discussed in the next section.

12.6.4 Extreme error to standard deviation conversion for different types of data set

According to the discussion above, the standard deviation of the errors resulting from the effect of topographic generalisation, SD(F) in this annotation, in the case of terrain profiles can be expressed as follows:

$$SD(F) = K\ Dx\ \tan A \qquad (12.22)$$

Where A is the slope angle and K is a proportional facter. The next step which needs to be carried out is to consider the value of K in different cases since more than one of the possible extreme errors may occur in a single case.

(1). The value of K for the DTM derived from composite data

For the composite data, the surface-specific points and surface-specific lines are all sampled, therefore, there is only one possible type of extreme error involved in the data points interpolated from the linear facets which are constructed from the composite data. In this case, the value of the possible extreme error is expressed by Equ.(12.19), thus the values of K, in this case, K_C, is

$$K_C = (1/4)\ (1/4) = 1/16 \qquad (12.23)$$

Therefore, the standard deviation of the corresponding distribution, $SD(F_C)$, can be estimated as follows:

$$SD(F_C) = E_C / 4 = Dx\ \tan A\ /16 \qquad (12.24)$$

(2). The value of K for the DTM derived from gridded data

For the gridded data, the geomorphological information is not included. In other words, the surface-specific points and surface-specific lines have not been included. In this case, all types of extreme values can possibly occur in the points interpolated from the DTM surface which are constructed from the gridded data. Therefore, the overall value of

the K in Equ.(12.22) should be the weighted average of all the different types of extreme errors which may occur.

Therefore, the value of K depends on the occurrence frequencies of the different types of extreme errors, i.e. the proportions of the slopes like those shown in Fig.12.5(a) and Fig.12.5(b). Therefore, the value of the K, in this case, K_r, can be expressed as follows:

$$K_r = U.P_u + V.P_v + T.P_t \qquad (12.25)$$

Where, P_t is the proportion of the grids covering the slopes with faults or discontinuities; P_v is the proportion of grids covering the slope with local minima or maxima (surface-specific points and those points lying along surface-specific lines); P_u is the proportion of the grids with normal slope; and $P_u = 1 - P_v - P_t$ in this case. If there are no faults or discontinuites in this area or if they have been measured and included in the data set, then the value of P_t is given as zero. Similarly, if points along form lines have been included, P_v should be given as zero. If P_v and P_t are both equal to zero, then it is the case that the DTM surface has been constructed from a composite data set.

By taking 4xSD as the limit of the maximum random errors, then the proportional factors for U and V can be expressed as follows:

V = 1/8; for those grids covering local minima or maxima; (12.26)
U = 1/16; for those grids covering normal (random) slopes.

The value of the proportional factor T is very difficult to tell since it depends the value of the E_b, which may vary from case to case. However, if the E_b is not too large, then 1/4 might be used as an approximate value. In practice, faults or discontinuities are included in most cases.

The next problem which arises is how to estimate the proportional factor for V and U. Of course, if V is known, then U=(1-V). Again, the estimation of the value of U is a very difficult matter to solve. For a very small area, there is no better method other than simply counting the numbers of points included for surface-specific lines in the area to be modelled. However, for a large area, the concept of "wavelength" may be used. According to the discussion conducted in Chapter 3, the value of such a wavelength can be estimated approximately as follows (Fig.12.6) (see p231a).

$$W = 2\ H\ \cot A \qquad (12.27a)$$

Where, H denotes the average relative height which may also vary from place to place; A is the slope angle; then W is the approximate value

of the so-called wavelength of the height variation present in the data set. The planimetric shape of a hill (expressed by contours) could be very different from place to place. Even for the same hill, the wavelength could be different if the profiles are taken along different directions and it is also not the same for different profiles even in the same direction. Therefore, an average value should be used. Therefore, if the value of $(H_{max}-H_{min})$ in the local area is used for the H in Equ.(12.27a), then the following equation might be appropriate:

$$W = (H_{max}-H_{min}) \cot A \qquad (12.27b)$$

After the estimated value of W has been obtained, the proportional factor - V - can then be estimated. As one can imagine, both the top and the bottom of a spatial variation will occur once over a single wavelength in one profile direction. Therefore, for a grid which has two profile directions perpendicular to each other, the occurrence frequency of the extreme values among the all acquired data points is as follows:

$$P_v = 4\,Dx / W \qquad (12.28)$$

By letting $P_t=0$ and substuting Equ.(12.28) into Equ.(12.25), then the following expression can be obtained:

$$\begin{aligned} K_r &= (1/16)(1- 4Dx/W) + (1/8)(4Dx/W) \qquad (12.29)\\ &= (1/16)(1+ 4Dx/W) \end{aligned}$$

Therefore, finally, the expression for the accuracy loss of the DTM data points interpolated from the DTM surface constructed from regular gridded data due to the generalisation effect, $SD(F_r)$, is as follows:

$$SD(Fr) = (1/16)(1 + 4Dx/W)\,Dx \tan A \qquad (12.30)$$

12.7 Overall accuracy of the digital terrain model surface

The accuracy loss of the digital terrain surface due to the two basic factors has been discussed individually in the previous sections (12.5 and 12.6). Next, the overall accuracy of the digital terrain model surface can be studied by integrating the accuracy losses due to these two factors. This section is an attempt to perform such an integration.

12.7.1 Accuracy of DTM points in the case of profiles

The case of the DTM points interpolated only along **profiles** needs first of all to be considered. In this case, according to the laws of error propagation, the value of the accuracy of DTMs, having regard to the

topographic generalisation effect and the errors in source data, can be written as follows:

$$\begin{aligned} VAR(Pr) &= VAR(P) + VAR(F) \\ &= (2/3)\ VAR(PMD) + (\ K\ Dx\ tanA\ )^2 \end{aligned} \qquad (12.31)$$

Where, VAR(Pr) denotes the variance of the DTM points in the case of profiles; VAR(P) is the accuracy loss due to the errors in the PMD; and VAR(F) is the accuracy loss due to the topographic generalisation effect.

A profile is a feature with only two dimensions. However, **a surface** is three-dimensional, therefore, the accuracy of the DTM points interpolated via profiles is not equal to that of a surface. The latter will be considered in the next paragraphs.

12.7.2 Accuracy of DTM points on a square-gridded cell modelled by a bilinear surface

The linear interpolation on a square-gridded bilinear surface has been discussed previously (Section 12.5.1) and the interpolation functions are expressed by Equs.(12.6) and (12.7). The accuracy of points H or G (Fig.12.3) can be expressed by Equ.(12.31). By taking into consideration the errors in the source data (points H and G) and the topographic generalisation effect, the accuracy of the points interpolated from the points H and G is therefore as follows:

$$VAR(sg) = (2/3)\ VAR(Pr) + VAR(F2) \qquad (12.32)$$

Where, the first term in the right side of the equation expresses the accuracy loss due to the errors propagated from points H and G, the accuracy of which is expressed by a variance - VAR(Pr); and the second term, VAR(F2), expresses the accuracy loss due to the topographic generalisation effect. In this case, VAR(F2) might be different from VAR(F) (see Equ.12.22). If the latter represents the accuracy loss due to the topographic generalisation effect in the profile direction, then the former is that in the direction perpendicular to the profiles; and vice versa. Since the average slope value is used to represent the characteristics of the terrain in any direction of the profiles, for the sake of convenience, they are assumed to be equal. Therefore,

$$\begin{aligned} VAR(sg) &= (2/3)(\ VAR(P) + VAR(F)\ ) + VAR(F) \\ &= (2/3)(\ 2/3\ VAR(PMD) + 5/3\ VAR(F) \\ &= 4/9\ VAR(PMD) + 5/3\ VAR(F) \\ &= 4/9\ VAR(PMD) + 5/3\ SD^2(F) \\ &= 4/9\ SD^2(PMD) + 5/3\ (\ K\ Dx\ tanA\ )^2 \end{aligned} \qquad (12.33)$$

Where, VAR(sg) denotes the variance of the DTM surface constructed from square-gridded data. The value of K varies with the data pattern which has been discussed previously (in Section 12.6.4). Therefore, for the **composite data** set (including the surface-specific points and lines), the accuracy of the DTM can be approximated by the following expression: -

$$\begin{aligned} VAR(sg)c &= 4/9\ VAR(PMD) + 5/3\ (K_c\ Dx\ tanA\ )^2 \\ &= 4/9\ VAR(PMD) + 5/3\ (\ 1/16\ Dx\ tanA\ )^2 \\ &= 4/9\ VAR(PMD) + 5/768\ (Dx\ tanA\ )^2 \end{aligned} \qquad (12.34)$$

On the other hand, for the **regular gridded data** only, the accuracy of the DTM is as follows:

$$\begin{aligned} VAR(sg)r &= 4/9\ VAR(PMD) + 5/3\ (K_r\ Dx\ tanA\ )^2 \\ &= 4/9\ VAR(PMD) + (5/768)((1 + 4Dx/W)\ Dx\ tanA\ )^2 \end{aligned} \qquad (12.35)$$

12.7.3 Accuracy of the DTM points interpolated using triangular facets

The interpolation of the DTM points on the continuous surface comprising a series of isosceles right-angled triangular facets has also been described (Section 12.5.1) and it has been noted that the interpolation procedure is very similar to that on sqaure-gridded bilinear surface. The only differences (see Fig.12.4) are as follows:

(a). Any side, DH or KF, parallel to either grid side is shorter than the square grid side;

(b). The length of the hypotenuse is 1.414 times that of the square grid side; and

(c). The slope of hypotenuse is on average 1.06 times the average value of the slope along either of the two grid directions (see Equ.(12.11)).

The average value of the accuracy loss at points H, G, and F in the hypotenuse resulted from the topographic generalisation effect is

$$SD(F_h) = K.\ 2^{1/2}\ Dx.\ (\ 2^{1/2} x\ 3/4)tanA = 3/2\ K.Dx.tanA \qquad (12.36)$$

Therefore, the accuracy of the points occuring along the hypotenuse considering both the errors in the source data and the generalisation effect is

$$\begin{aligned} VAR(hy) &= 2/3\ VAR(PMD) + VAR(F_h) \\ &= 2/3\ VAR(PMD) + 9/4\ (K\ Dx\ tanA\ )^2 \end{aligned} \qquad (12.37)$$

Next an estimation of the accuracy of the DTM points derived from the triangular facets needs to be made. Taking the interpolation along HD (Fig.12.4) as an example, the accuracy of the DTM points interpolated from such a surface can be expressed as follows:

$$VAR(tg) = 1/3\ VAR(pr) + 1/3\ VAR(hy) + (K.HD.tanA)^2 \qquad (12.38)$$

Where, the terms in the right side of the equation are as follows: the first term represents the accuracy loss due to the errors in the point D (i.e. along the grid side); the second term expresses the accuracy loss due to the errors in point H (i.e. along the hypotenuse); and the third term represents the accuracy loss resulting from the generalisation effect. In this case, considering the length of line DH as a variable, it varies from the position to position. Therefore, the average value for $(K\ HD\ tanA)^2$ should be used and it can be calculated as follows:

$$\begin{aligned}(K\ HD\ tanA)^2 &= (K\ tanA)^2\ (1/Dx)\int_0^{Dx} Y^2\ dY \\ &= (1/3)(K.Dx.tanA)^2\end{aligned} \qquad (12.39)$$

By substituting Equs.(12.39), (12.37) and (12.31) into (12.38), the following can be obtained:

$$\begin{aligned}VAR(tg) = &\ (1/3)\ (\ 2/3\ VAR(PMD) + (K.Dx.tanA)^2\ ) \\ &+ (1/3)\ (\ 2/3\ VAR(PMD) + 9/4\ (K.Dx.tanA)^2\ ) \\ &+ (1/3)\ (K.Dx.tanA)^2 \\ = &\ (4/9)\ VAR(PMD) + (17/12)\ (K.Dx.tanA)^2 \\ = &\ (4/9)\ SD^2(PMD) + (17/12)\ (K.Dx.tanA)^2\end{aligned} \qquad (12.40)$$

Following a similar procedure to that discussed for the bilinear surfaces, the accuracy of the DTM points on the triangular facets constructed from **composite data** may be expressed as follows:

$$\begin{aligned}VAR(tg)c &= (4/9)\ VAR(PMD) + (17/12)\ (K_c.Dx.tanA)^2 \\ &= (4/9)\ VAR(PMD) + (17/3072)\ (Dx.tanA)^2\end{aligned} \qquad (12.41)$$

However, for the triangular surface constructed from only regular **gridded data**, the accuracy of the DTM points is as follows:

$$\begin{aligned}VAR(tg)r &= 4/9\ VAR(PMD) + 5/3\ (K_r\ Dx\ tanA)^2 \\ &= 4/9\ VAR(PMD) + 17/3072\ ((1+4Dx/W)\ Dx\ tanA)^2\end{aligned} \qquad (12.42)$$

12.7.4 Summary of the important formulae

For the sake of convenience in comparison, the **important formulae** developed in this section are summarized as follows:

i). Accuracy of DTM data interpolated along **profiles:**

$$VAR(Pr) = (2/3)\ VAR(PMD) + (\ K\ Dx\ tanA\ )^2 \qquad (12.31)$$

ii). Accuracy of DTM data interpolated from the **hybrid of bilinear surfaces and triangular facets** from **composite data** (i.e. in the case that surface-specific points and lines have been included in the source data):

$$VAR(sg)c = 4/9\ VAR(PMD) + 5/768\ (Dx\ tanA\ )^2 \qquad (12.34)$$

iii). Accuracy of DTM data interpolated fron the **bilinear surfaces:**

$$VAR(sg)r = 4/9\ VAR(PMD) + 5/768\ ((1 + 4Dx/W)Dx\ tanA\ )^2 \qquad (12.35)$$

where, $$W = (H_{max}-H_{min})\ cotA \qquad (12.27b)$$

iv). Accuracy of the DTM points on the **triangular facets** constructed from **composite data** is as follows:

$$VAR(tg)c = (4/9)\ VAR(PMD) + (17/3072)\ (Dx.tanA)^2 \qquad (12.41)$$

v). Accuracy of DTM data interpolated from the **triangular facets** which are constructed from the **gridded data** set:

$$VAR(tg)r = 4/9\ VAR(PMD) + 17/3072\ ((1+4Dx/W)\ Dx\ tanA\ )^2 \qquad (12.42)$$

12.8 Evaluation of various new theoretical models

After the mathematical models of DTM accuracy have been established for various terrain situation, their "goodness" requires to be judged. In this section, first of all, these models will be compared with the experimental results presented previously in this thesis; then they will be evaluated from a theoretical standpoint.

12.8.1 Experimental validation of new accuracy models

Some experimental test results have been presented in Chapter 9 and earlier in this chapter. In this section, an attempt is made to use these test results to validate the mathematical models which have just been developed above.

(1). Comparison of the test results obtained from the Munich ISPRS data with the predictions by the new accuracy models

In Chapter 9, it was intended to use an empirical model to fit the

experimental data. It was found that this seems to work well for the accuracy of DTM data derived from composite data sets, but not for gridded data sets. It can also be found clearly that the theoretical model for composite data sets which has been just described above is exactly the same as that used in Chapter 9. However, the model used for gridded data sets is different because the occurrence frequency (or probability if the data set is very large) of the grids covering local minima or maxima (points along form lines) varies with sampling interval.

Fig.12.7 shows in diagrammatic form the comparison of theoretical models with the experimental data. It can be found that they fit rather well. The worst case concerns the gridded data sets for Spitze. This is still within the limits of expectation since it was found from the test results given in Chapter 9 that the occurrence frequencies of gross errors for these data sets are very high due to the existence of faults or discontinuites. The parameters used for the theoretical models which have produced Fig.12.7 are as follows:

The **slope angles** are estimated from the contours. The values used for the Uppland, Sohnstetten and Spitze areas are 6, 15 and 7 degrees, respectively. The **accuracies of the source data** in terms of the VAR(PMD) values for different data sets are also estimated in Chapter 9 as follows: 0.67m for the Uppland set, 0.16m for the Sohnstetten set and 0.08m for the Spitze set.

For the accuracy model for the DTM surface constructed from only gridded data, the so-called **terrain wavelength** needs also to be estimated. As has already been discussed in Chapter 9, for the Sohnstetten area, the test area is actually located along a valley in a mountainous area. The width of such a profile (for the area with check points because only this area has been tested) is about 214m. This value is considered as the wavelength for this area and can be written as W=214m. Similarly, the value of W for the Spitze area is given as 300m. The value of W for Uppland has been estimated according to the discussions in this chapter as 470m. Also for the Spitze area, two faults or discontinuites exist. The height variations encountered in the different segments of the discontinuites are different, varying from 3m to 0.5m. The average height variation has been estimated (from the contours) to be 1.25m, thus E_b=1.25m. P_t for these discontinuites is 4Dx/W. According to these estimates, the curve lines for the composite data sets are also presented in Fig.12.7 for comparison with experimenal data.

The purpose of presenting such descriptions for the parameter estimation is to give readers some idea about the adequacy of these parameters, and thus form some impression of the adequacy of these

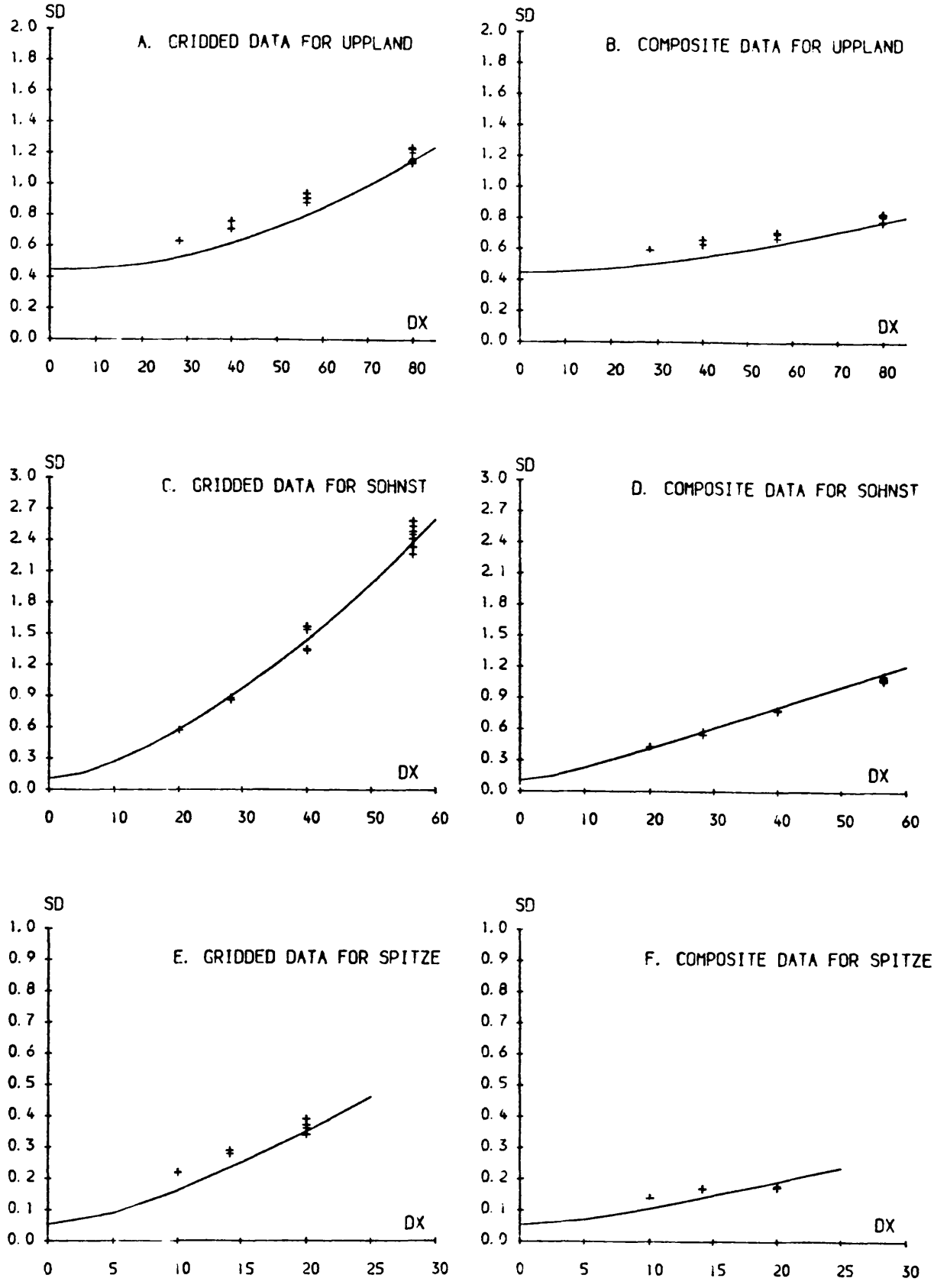


FIG. 12.7 VARIATION OF DTM ACCURACY WITH SAMPLING INTERVAL

(+ STANDS FOR TEST RESULTS, SOLID LINE FOR THEORETICAL MODEL)

theoretical models.

In order to provide more complete information about the accuracy of the predictions produced by these new models, some figures corresponding to Fig.12.7 have also been given in Table 12.8. It can be seen clearly that vast majority of the differences are below 0.1m. It goes without saying, these results are indeed very good.

Table 12.8 Comparison of accuracy predicted by new models with test results obtained from the Munich data sets

Test	Grid	Gridded Data			Composite Data		
Area	Interval	Predicted	Tested	Differ.	Predicted	Tested	Differ.
Upp-land	28.28m	0.53m	0.63m	-0.10m	0.50m	0.59m	-0.09m
	40m	0.62m	0.76m	-0.14m	0.55m	0.66m	-0.11m
	56.56m	0.80m	0.93m	-0.13m	0.63m	0.70m	-0.07m
	80m	1.16m	1.18m	-0.02m	0.78m	0.80m	-0.02m
Sohns-tetten	20m	0.58m	0.56m	0.02m	0.42m	0.43m	-0.01m
	28.28m	0.90m	0.87m	0.03m	0.58m	0.56m	0.02m
	40m	1.44m	1.45m	-0.01m	0.81m	0.78m	0.03m
	56.56m	2.38m	2.40m	-0.02m	1.15m	1.08m	0.07m
Spitze	10m	0.16m	0.21m	-0.05m	0.11m	0.16m	-0.05m
	14.14m	0.23m	0.28m	-0.05m	0.14m	0.17m	-0.03m
	20m	0.35m	0.35m	0.00m	0.19m	0.18m	0.01m

(2). Comparison of test results obtained from the Stuttgart data with the prediction by the new accuracy models

Earlier in this chapter, the results obtained from the data sets measured at the University of Stuttgart have been used to evaluate the existing mathematical models of DTM accuracy based on the summation of Fourier spectra over their high-frequency parts. These test results will now also be used to evaluate the mathematical models which have just been described above to see how well these models fit the results from the Stuttgart data. The results are shown in Table 12.9.

In this estimation, the slope angle for Drivdalen was taken as 40°. The width of the area covered by the check points, 1,200m, has been used as the wavelength for this test. It can be seen clearly that the predictions produced by these accuracy models developed in this project are very close to the test results. Indeed, it may be said that the predictions are very good.

Table 12.9 Comparison of test results from Stuttgart data with the values predicted by new accuracy models

Test Area	Grid data			Composite data		
	Predicted	Tested	Differ.	Predicted	Tested	Differ.
Spitze	0.25 m	0.31 m	-0.06m	0.16 m	0.20 m	-0.04m
Sohnstetten	0.40 m	0.46 m	-0.06m	0.32 m	0.35 m	-0.03m
Drivdalen	1.60 m	1.57 m	0.03m	1.57 m	1.47 m	0.10m

12.8.2 Theoretical evaluation of the new DTM accuracy models

As has been discussed in Chapter 2, seven characteristics of mathematical models can be used as the **standards** to judge their "goodness". These are their resulting accuracy, descriptive realism, precision, robustness, generality, fruitfulness and simplicity. Therefore, these seven standards will be applied to the mathematical models of the accuracy of digital terrain model surfaces which have just been developed in this chapter.

The **accuracy** of the new models will first of all be considered. It has been defined in Chapter 2 that a model is said to be **accurate** if the output of the model (i.e. the answer it gives) is correct or very near to being correct. The comparison of the new models with the ISPRS experimental data shows that the predicted results are very near to the experimental data. Therefore, at least on the basis of these limited tests, it can be said that these new models are accurate.

The **descriptive realism** of the new accuracy models will then be considered. It has been defined that a model is said to be **descriptively realistic** if it is based on assumptions which are correct. In the development of the accuracy models in this study, a basic assumption which has been used that the two extreme values of a distribution are about 4 times the standard deviation of the distribution (distant from the mean of the distribution) in the case of digital terrain models. As has been discussed previously (Section 12.6.3), this assumption seems appropriate both in theory and practice. Therefore, it can be said that these new models are descriptively realistic.

It has also been defined in Chapter 2 that a model is said to be **precise** if its predictions are definite numbers (or other definite kinds of mathematical entities, such as functions, geometric figures, etc.). Obviously, all the accuracy models developed in this study are

expressed by mathematical functions which can also be presented in a graphical form. Therefore, by this definition, they are precise models.

According to the definition given in Chapter 2, a model is said to be **robust** if it is relatively immune to errors in the input data. Considering the particular cases of the new accuracy models, the input parameters are the sampling interval, the slope angle, the accuracy of source data and perhaps the wavelength. In these models, the sampling interval plays the most important role. However, in any case, the sampling interval will be very precise. As far as the other parameters are concerned, a small error in the input will merely produce a small error in the output. Therefore, in a certain sense, these new models can be said to be relatively robust.

It has been defined in Chapter 2 that a model is said to be **general** if it applies to a wide variety of situations. In some ways, it can be said that the new mathematical models are **not general** since they apply only to the limited specific cases - employing (i) either gridded data or composite data; (ii) a continuous surface comprising a series of linear facets which are constructed directly from the measured data (i.e. without the pre-processing of a random-to-grid interpolation); and (iii) a relatively small sampling interval.

The last **limitation** must be emphasized here, i.e. these models will apply to only these cases where the sampling interval is smaller than the so-called wavelength. This limitation is obvious since the values of the maximum errors cannot increase any more when the sampling interval reaches the value equal to the wavelength of the terrain variation.

The **fruitfulness** of a model is also very important. It has been defined that a model is said to be **fruitful** if either its conclusions are useful or it inspires and/or points the way to other good models. Obviously, the accuracy models developed in this study would appear to be useful for application in DTM practice. They also point the way to the definition of an optimum sampling interval which will be discussed later in Chapter 14.

Simplicity is also an important standard for a mathematical model, especially for the purpose of comparison. A model may be regarded as **simple** if it comprises only a small number of parameters, each of which has some specific or special meanings. It is clearly the case that the accuracy models developed in this study have been derived in an intuitive way and no complicated mathematics are needed at all. Furthermore, each model consists of only a few parameters, each of which has its own physical meaning. Finally, it should be said that the mathematical expressions of these models are extremely simple and are

in a form which is similar to that of the traditional expression of map accuracy. It might be believed (at least by the present author) that these models are the simplest ones which can ever be obtained. Therefore, if this is indeed correct, it is obvious that these models also satisfy the standard of simplicity.

From the discussion conducted above, it would seem that these models do satisfy the **six standards** out of the total of seven. The exception is that of the generality. Actually, generality is not a serious problem in this particular case since the gridded data and composite data are the two types of DTM source data which are most widely used in DTM practice and since the continuous surface comprising a series of linear facets is, by general consent, the least misleading surface which can be used in most cases. Therefore, it can be understood that these models may be placed in a fairly high rank in terms of "goodness".

12.9 Concluding remarks

In this chapter, existing mathematical models of the DTM accuracy have been evaluated and two main types of mathematical model describing the accuracy of the DTMs derived from photogrammetrically measured data sets have been described. One of the models is for composite data, while the other is for regularly gridded data - both of which are commonly used in DTM practice. These mathematical models have also been validated with the experimental results which have been presented in Chapter 9 and have also been discussed using the seven standards given in Chapter 2. As a result, it can be concluded that these models are reasonably apt and fairly comprehensive models.

It can also be found that these formulae expressing the accuracy models are so simple that they can also be expressed in a form similar to that of Koppe formulae (see Chapter 6) which have been widely used in the mapping communities in Central European countries such as Germany, Austria and Switzerland to express the accuracy of a contour map. The expressions corresponding to the form of the Koppe formulae might be expressed as follows:-

i). Accuracy of DTM data interpolated from the hybrid of **bilinear surfaces and triangular facets** from **composite data** (i.e. in the case that surface-specific points and lines are included in the source data):

$$SD(sg)c = (2/3)\ SD(PMD) + 0.081\ Dx\ \tan A \qquad (12.43)$$

ii). Accuracy of DTM data interpolated from **bilinear surfaces:**

$$SD(sg)r = (2/3)\ SD(PMD) + 0.081\ (1 + 4Dx/W)\ Dx\ tanA \qquad (12.44)$$

iii). Accuracy of the DTM points on the **triangular facets** constructed from **composite data** is as follows:

$$SD(tg)c = (2/3)\ SD(PMD) + 0.075\ Dx.tanA \qquad (12.45)$$

iv). Accuracy of DTM data interpolated from the **triangular** facets which are constructed from a **gridded data** set:

$$SD(tg)r = (2/3)\ SD(PMD) + 0.075\ (1+4Dx/W)\ Dx\ tanA \qquad (12.46)$$

So far, the matter of the accuracy of the DTMs derived from photogrammetric data has been investigated in some detail. However, the problem of the accuracy of the DTMs derived from digital contour data has not as yet been touched upon. The next chapter is an attempt to conduct some useful investigations into this specific topic.

Chapter 13

Accuracy of Digital Terrain Models Derived from Digital Contour Data

Chapter Thirteen

Accuracy of Digital Terrain Models Derived from Digital Contour Data

13.1 Introduction

As has been discussed in Chapter 4, based on practical considerations, aerial photography and the contour lines on existing topographic maps are the two main sources for modelling large areas of the Earth's surface. Therefore, the accuracies of the DTMs derived from both of these sources have been investigated in this project. Some experimental tests on the accuracies of the DTMs derived from the data acquired by photogrammetric sampling have already been described in previous chapters (9 and 10) and a family of mathematical models predicting the accuracy of the DTMs derived from photogrammetrically measured data has been presented in Chapter 12. However, the accuracies of DTMs derived from digital contour data have not been discussed till now. This chapter is an attempt to deal with this subject.

As has been pointed out in Chapter 6, the term "digital contour data" includes both the digital data digitised from existing contour maps and the digital data recorded directly from photogrammetric contouring. In this project, an experimental test on each of these two types of contour data has been carried out and will be reported on in this chapter. Actually, the first test was an attempt to investigate the accuracy of the DTMs derived from contour data related to different slope angles of the terrain while the second test was designed to investigate the possible improvement in the accuracy of the DTMs derived from digital contour data through the addition of feature-specific data.

13.2 A study of the DTM produced from an OS 1:63,360 scale topographic map

A brief description of OS topographic map series has been given in Chapter 6. In this section, an experimental test on the accuracy of the DTM derived from a 1:63,360 scale map will be presented.

13.2.1 Introduction

Existing topographic maps are one of the main sources for the establishment of a small-scale topographic data base and for digital terrain modelling on a regional basis. In Great Britain, the 1:50,000 scale topographic map series was recommended to be the principal data source for such purposes by the Ordnance Survey (OS) Review Committee (Haywood, 1981). Indeed, McMaster et al (1986) reported that the OS was

concentrating on "the creation of a data base from its 1:50,000 scale digital maps,..." and that the OS was also "examining 1:50,000 digital data produced by other agencies with a view to using this in the data base and had started digitising one 1:50,000 scale sheet for assessment by potential users". Therefore, a study of the accuracy of the DTMs which could be derived from topographic maps at such a scale is of practical interest.

A preliminary test carried out by Shearer and one of his students (Lowthian, 1986; Shearer, 1987) in this Department for the production of a DTM from this scale of OS topographic map gave very disappointing results. The resultant root mean square error (RMSE) was about $\pm$6.75m for a low mountain area.

So it seemed particularly interesting to test the accuracy of DTM from an Ordnance Survey (OS) 1:63,360 (one inch to mile) scale topographic map, not only because these two scales are quite close but, more importantly, because the OS 1:50,000 scale series is the successor to the 1:63,360 series after metrication. "The 1:50,000 First Series sheets are basically an enlargement of the material of the Seventh Series 1:63,360 sheets". The contours are straight conversions from feet to the nearest metre. As Harley (1975) mentioned, "the actual contour interval of the 1:63,360 scale series - 50ft - remains the same". Such a test may also facilitate the feasibility study for creating a small topographic data base.

In addition to these reasons mentioned above, a large area of digital contour data digitised from the 1:63,360 scale OS topographic maps in East Fife which has been used for testing the rectification of remotely sensed image data was available in this Department. For these reasons, a test of the accuracy of a DTM derived from one of these 1:63,360 scale topographic maps was carried out.

13.2.2 Test area

The digitized area covers a complete sheet - Sheet 56. The coordinate range of the test area is approximately as follows (see Fig.13.1):

Easting:	325,000 m ----- 360,000 m
Northing:	700,000 m ----- 725,000 m

The relief in this area belongs to three categories as follows:

i). high hills covering an elevation range between 700 and 2,000ft (210 to 600m);

ii). low hills lying between 350 and 700ft (105 to 210m); and

iii). a considerable area of lowland and shallow valleys.

Slope angle ranges from 1^{o} to 30^{o}. The contour lines in the test area are shown in Fig.13.1.

Because of the large variation in relief and slope angle in this area and the capacity limitations of the terrain modelling package (PANACEA) used for data processing, the test area was divided into 8 sub-areas. The relief parameters are shown in Table 13.1, where the slope angles have been estimated from the contour map itself and they should be regarded only as approximate values.

Table 13.1 Relief parameters for Sheet 56

Parameters	A1	A2	A3	A4	A5	A6	A7	A8
Slope range	6-30	6-30	5-25	3-25	3-25	3-15	1-13	3-13
Typical slope	24	26	15	10	13	7	4	6
Contour range	50-550	50-700	50-500	100-1100	50-950	50-650	50-600	50-650

N.B. The slopes are in a unit of "degrees" and the contours are given in "feet".

13.2.3 The digitised contour data

With the digitised contour data, two aspects are of main concern. One is the quality of the contour data included on the maps and the other is the "goodness" of the digitisation.

Table 13.2 Statistics of Digital Contour Data

Values on Ground		Values on Map	
D_{min}	1.6 m	d_{min}	0.03mm
D_{max}	85.7 m	d_{max}	1.35mm
D_{av}	40.0 m	d_{av}	0.63mm
No. of Pts	about 66,200	density	87pts/km^2

N.B. D_{min}, D_{max} and D_{av} denote the shortest, longest, and average distances on the ground; and
d_{min}, d_{max} and d_{av} denote the corresponding distances at map scale.

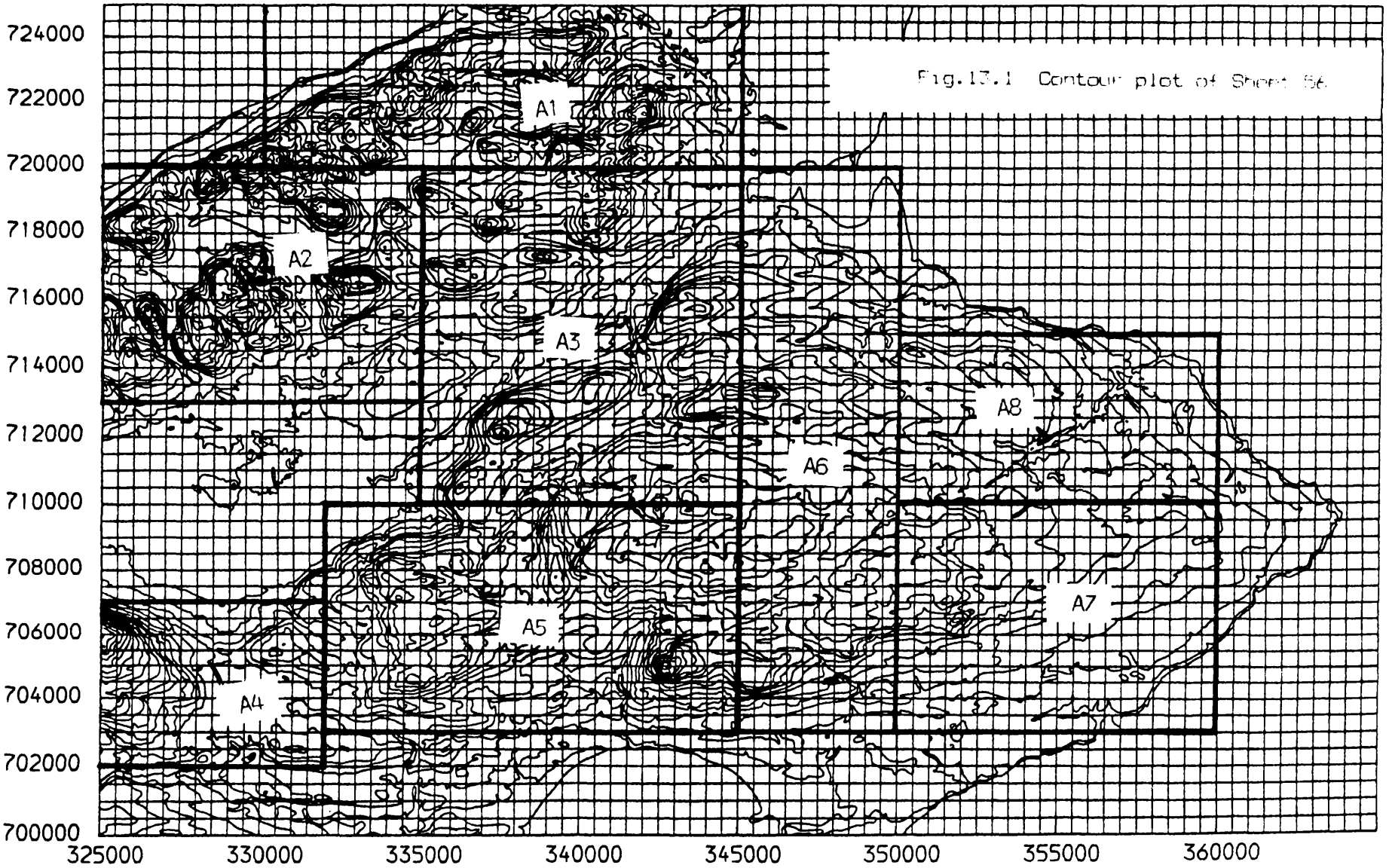


Fig.13.1 Contour plot of Short Se

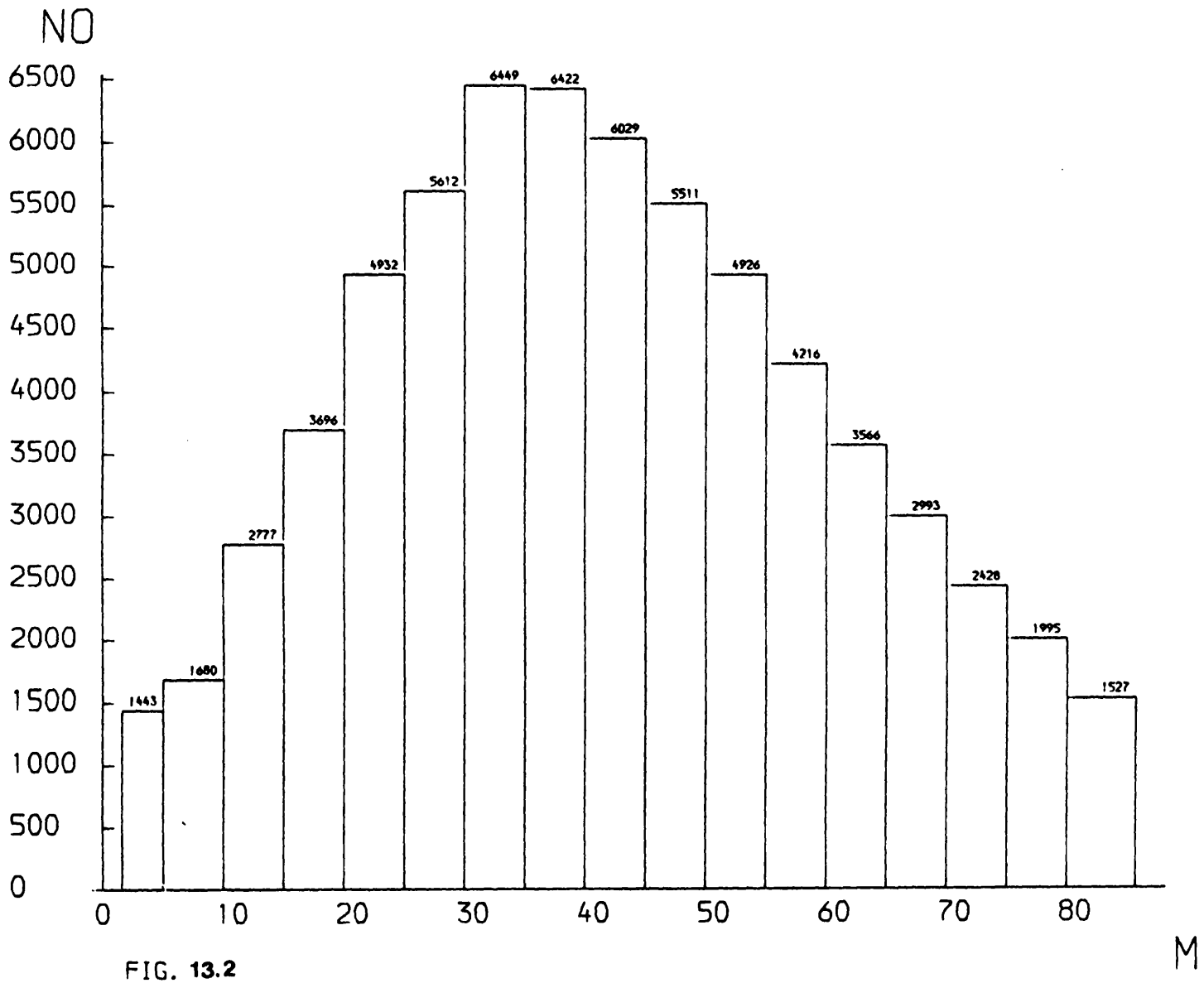


FIG. **13.2**

DISTRIBUTION OF DISTANCES BETWEEN TWO DIGITIZED POINTS

The vertical contour interval on the 1:63,360 scale maps is 50ft (about 15m). The contour lines were digitized manually on a Hewlett-Packard digitising system using stream mode. The distribution of the distances (chords) between pairs of digitised points is shown as a histogram in Fig.13.2. The corresponding statistics regarding these distances and the number and density of the points which are included in the digitised contour data set are also given in Table 13.2. From Fig.13.2, it can also be seen that about 86 per cent of the distances between pairs of digitized points are smaller than 1 mm at map scale which is equivalent to about 65m on the ground.

According to the discussion in Chapter 6, the accuracy of the digitised contour data for this area will be about 2m on average if the contours were compiled from the manuscripts measured from a stereo-model. However, the actual accuracy of this set of digital contour data should be lower (worse) than this value since the contours on OS 1:63,630 scale maps were derived from the older series of OS 1:10,560 scale maps.

13.2.4 The check points

As discussed before, three main aspects of the check points are of concern. One of them is their geometric accuracy; another is the number of check points which are available for checking; and the third is the distribution of those points.

As far as the **accuracy** is concerned, photogrammetric data is desirable for checking the accuracy of a DTM derived from contour maps. However, as already noted in the discussion of the accuracy of the DTM data derived from space photography in Chapter 11, it is impractical to use large scale photographs for the measurement of check points for so large an area. Therefore, attention was concentrated on the availability of larger scale contour maps.

As discussed in Chapter 6, the largest scale at which the OS measures and compiles contour maps is 1:10,000. However, the corresponding maps for this area were not available in the Department. Therefore, the OS (Pathfinder Series) 1:25,000 scale topographic maps were considered as an alternative. In other words, points digitised from the contours on such maps were to be used as check points for this test.

The 1:25,000 scale maps for this area are the recently published sheets of the OS Pathfinder Series. The sheets of this Series have in fact been compiled directly by simple photographic reduction from the 1:10,000 or 1:10,560 Series published in the 1970s and 1980s. The contour interval is 5m and the contours themselves have been measured

photogrammetrically. According to the map specifications discussed in Chapter 6, points interpolated from such contours will have an RMSE of about ±1.7m. The accuracy of the points actually lying on the contours may be a little higher than this value. On the other hand, according to the map specifications published by the OS, the accuracy (RMSE value) of points interpolated from the contours on 1:63,360 scale topographic map with 50 feet (about 15m) contour interval is about ±5m even in flat areas. Therefore, the points digitised along the contours on the 1:25,000 scale maps were considered acceptable to be used as check points for this test.

Another important concern with check points for the testing of DTM accuracy is their **distribution.** In this test, only those points occurring at certain selected height values have been used and these points may be not representative in some sense. Nevertheless, an attempt had also been made to select the check points as randomly as possible, and also as evenly distributed as possible. Fig.13.3 is an example of the point distribution.

Regarding the **number** of check points used, as discussed previously (in Chapter 11), the use of more check points may lead to a more reliable test result. However, when the accuracy of the check points themselves is relatively low, not much improvement in the reliability of the DTM accuracy estimates can be achieved even if a large number of check points are used. Therefore, it was decided not to use as many points as had been used in the ISPRS DTM test. The numbers of check points used in this test are shown in Table 13.3.

13.2.5 Test results and the analysis

Again, the PANACEA system (a triangulation-based DTM package) was used for the DTM generation. The digitised contour data was, first of all, triangulated; then the continuous surface comprising a series of contiguous linear facets was constructed from the triangular network. Finally, the DTM points were linearly interpolated from the triangulated facets. Fig.13.4 shows an example of the triangular network which has been constructed from the digital contour data of area A7.

The same statistical parameters as those used in the previous tests were computed and the test results are shown in Table 13.3. Also as was done before, the occurrence frequencies of the large residuals were also recorded and are given in Table 13.4.

From Table 13.3, it can be seen that the extreme values of the residuals are more or less equal to the contour interval although a few exceeded this value. The RMSE and SD values obtained from this test

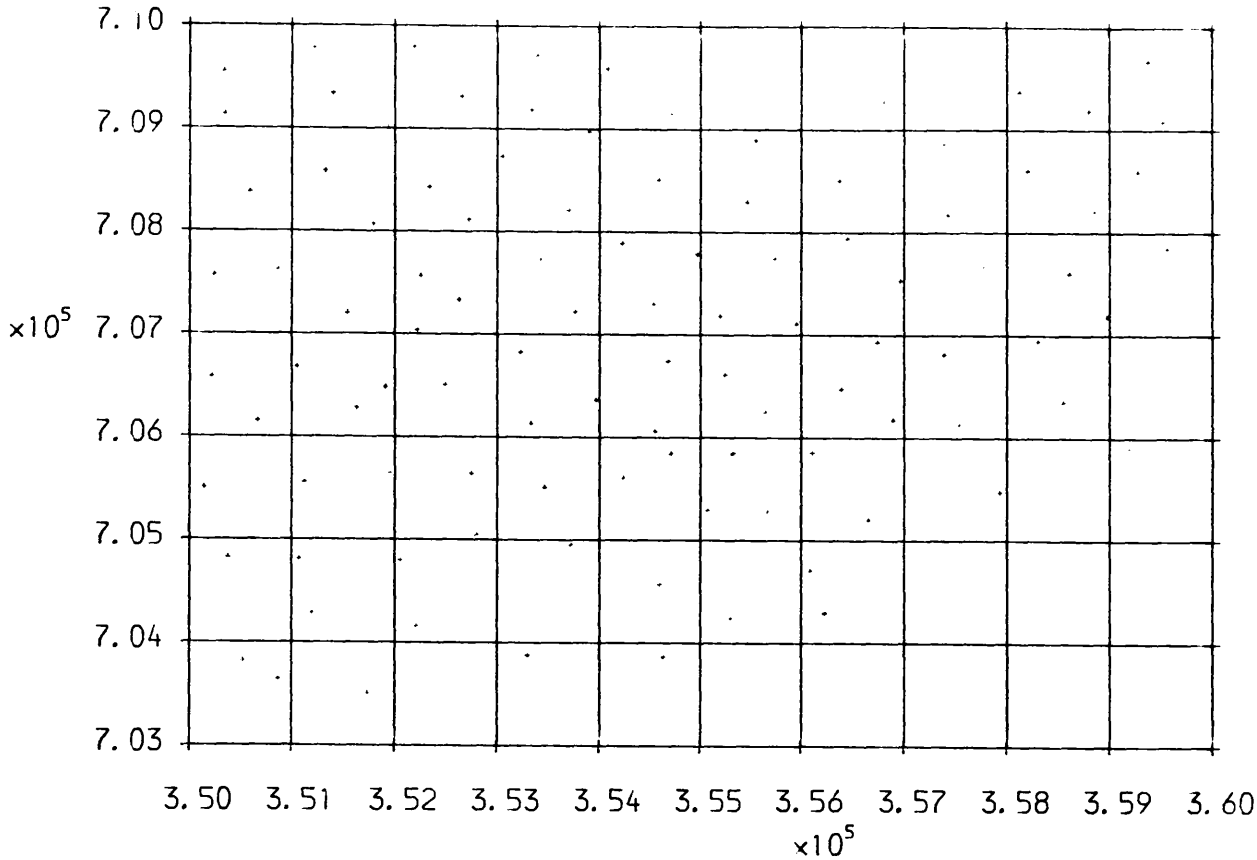


Fig.13.3 Distribution of check points for A7

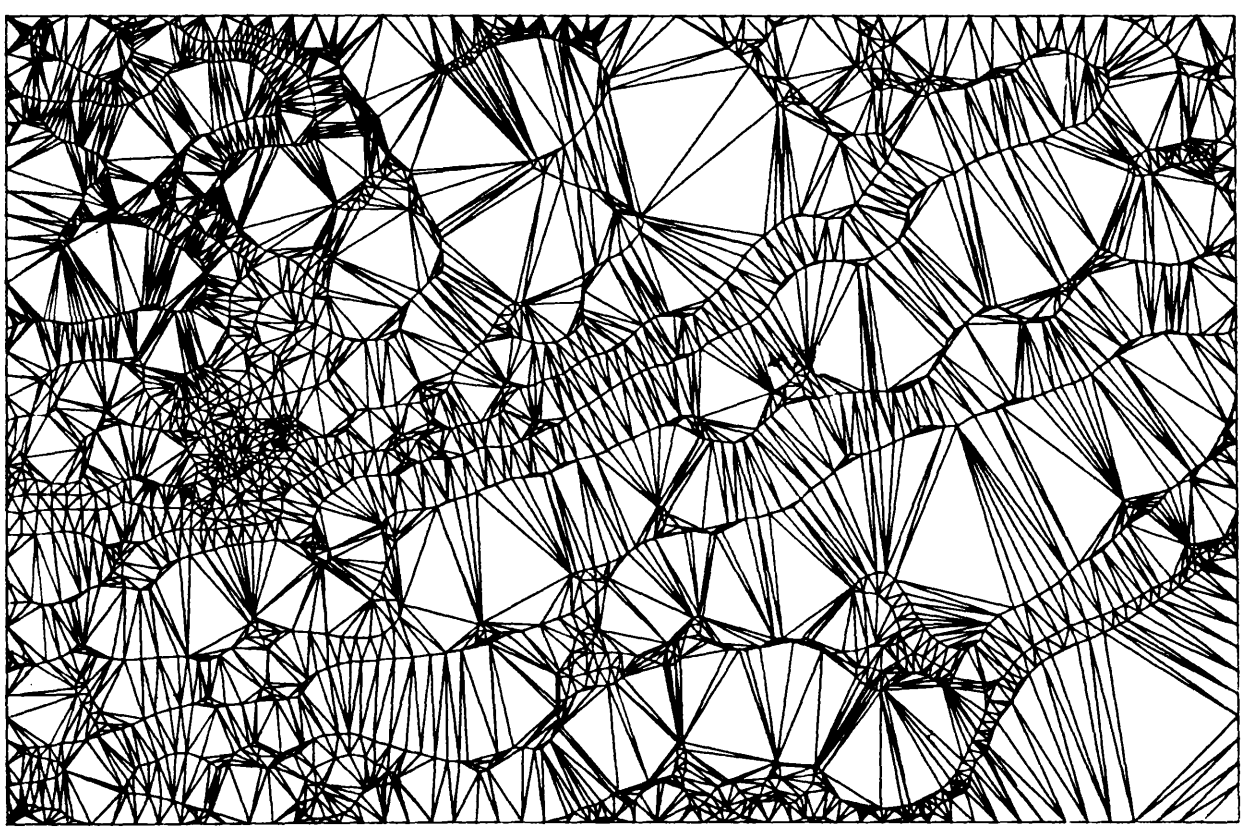


Fig.13.4 Triangular network for Area A7

show that the same accuracy figures as those given by the map specification may be obtained for the DTM derived from the digitised contour data. The overall RMSE value is ±4.4m, which is even a little better than the expected value: CI/3 = 15/3 = 5.0m. Of course, it is also clear that the RMSE values vary greatly with the actual terrain features.In the two hilliest areas with greatest slope values (i.e. areas A1 and A2), the RMSE values exceeded 5.0m, while in the four flatter areas with lowest average slope values (i.e. areas A4, A5, A7, and A8), the RMSE values were below ±4.0m.

Table 13.3 Test results for Sheet 56

	A1	A2	A3	A4	A5	A6	A7	A8	Overall
No.	117	94	168	74	137	178	104	88	960
E_{max}	16.2	16.6	21.7	7.1	6.9	10.2	6.2	8.3	21.7
$-E_{max}$	-17.2	-15.1	-16.7	-14.9	-22.2	-12.8	-11.6	-10.0	-22.2
RMSE	5.6	5.9	4.4	3.9	4.1	3.9	3.1	3.9	4.4
Mean	0.2	-1.0	-0.2	-1.5	-0.6	0.5	-0.2	-0.9	-0.3
SD	5.6	5.8	4.4	3.7	4.0	3.8	3.1	3.8	4.4

N.B. No. denotes the sample size (number of check points); RMSE denotes root means square error; and SD standard error; and the symbol "±" before RMSE and SD has been omitted simply for convenience. E_{max} and $-E_{max}$ denote the maximum and minimum errors; The unit for accuracy figures used in this table is the metre.

Table 13.4 Occurrence frequency of large errors

Area	Number			Frequency		
	>2xSD	>3xSD	>4xSD	>2xSD	>3xSD	>4xSD
A1	6	1	0	5.13 %	0.85%	0.0%
A2	5	0	0	5.32 %	0.0%	0.0%
A3	11	3	1	6.55 %	1.78%	0.6%
A4	3	1	0	4.05 %	1.35%	0.0%
A5	3	2	2	2.19 %	1.46%	1.46%
A6	5	2	0	2.81 %	1.12%	0.0%
A7	4	1	0	3.85 %	0.96%	0.0%
A8	6	0	0	6.82 %	0.0%	0.0%

The **occurrence frequencies** of large residuals are shown in Table 13.4. From this table, it can be seen that large errors (those larger than 3xSD) which might often be considered to be gross errors do occur in six out of the eight test sub-areas, with a small but discernible frequency. The errors larger than 4xSD did not occur in any of the areas with the lowest average slope values (areas A4, A6, A7, and A8) nor in the two hilliest areas (A1 and A2).

Since, as noted above, there is a **trend** visible in Table 13.3 for the RMSE or SD values to rise with an increase in the slope (and relief). it was decided to carry out a more detailed analysis in order to have a deeper insight into this trend.

Using the same empirical model and procedure as that used in Chapter 9, a similar regression analysis was carried out. The mathematical expression was rewritten as follows:

$$VAR(DTM) = a^2 + (b \tan A)^2 \qquad (13.1)$$

The coefficients "a" and "b" can be obtained by means of a regression analysis using the experimental data. "A" denotes the slope angle. If the so-called typical slope angles for those sub-areas listed in Table 13.2 are used for the regression analysis, then the resultant values for "a" and "b" are 3.5 and 9.86, respectively. The value of the correlation coefficient obtained for this regression, r, is 0.987. And so, the final expression becomes:

$$VAR(DTM) = 12.25 + (9.86 \tan A)^2 \qquad (13.2)$$

VAR(DTM) in Equs.(13.1) and (13.2) denotes the variance of the DTM and the standard error can be obtained by taking the square root of the value of VAR(DTM). The diagrammatic representation is shown in Fig.13.5.

Equ.(13.2) can only be an approximate one because the slope angles listed in Table 13.2 are not accurate values. Nevertheless, it does provide some useful information about the relationship between the DTM accuracy and the slope angle of the terrain. Indeed, Equ.(13.2) is very similar to the relationship actually set out in the map specifications discussed in Chapter 6.

13.2.6 Discussion

From this limited test, the following conclusions may be made:

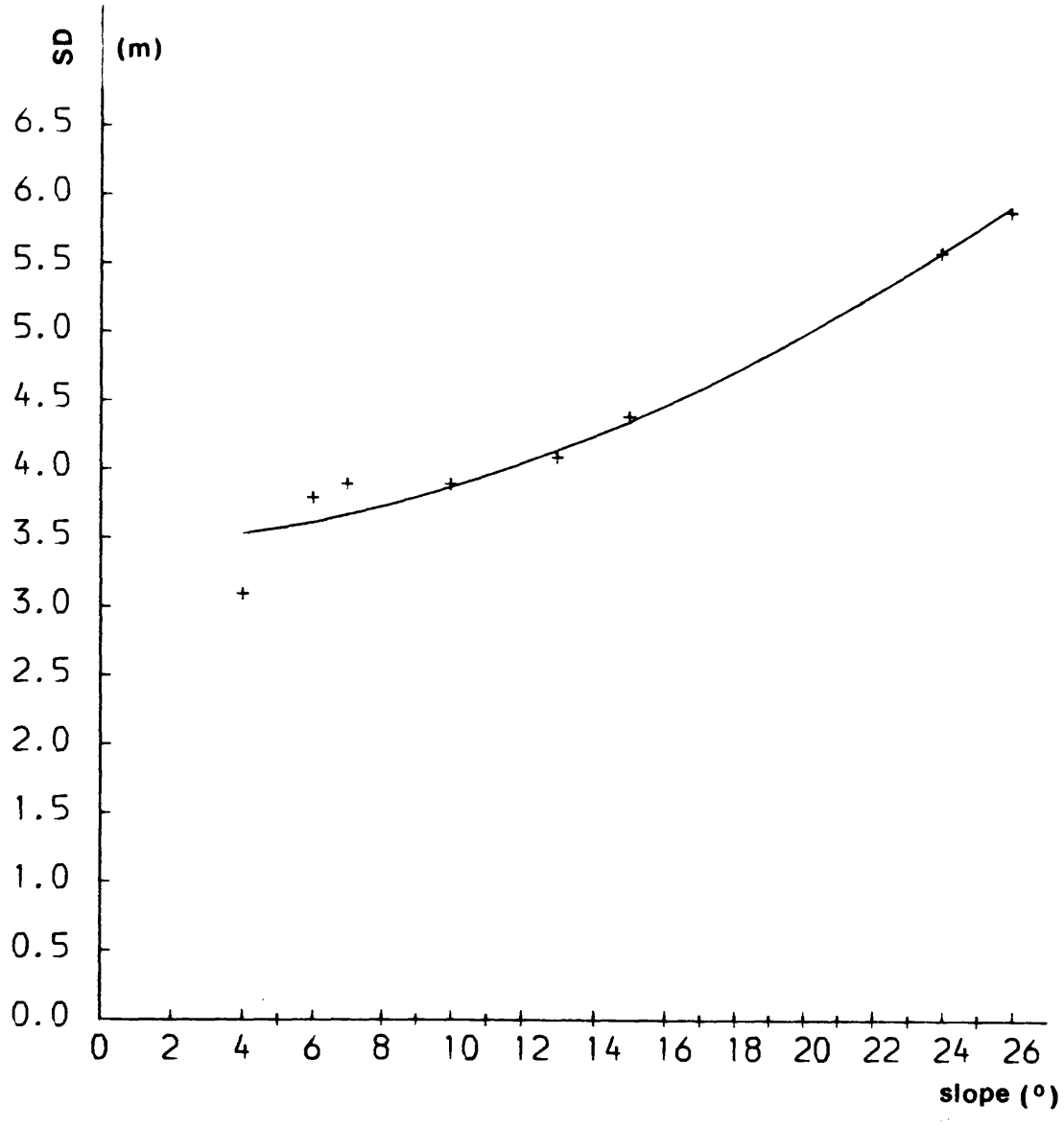


Fig.13.5 Accuracy of DTMs from digital contour data (OS 1:63,360 scale map) with slope

i). The height accuracy of the DTM data derived from contours is highly correlated with the slope angles occuring in the test area. The correlation coefficient is very high (0.987 for this example) for the model expressed by Equ.(13.1).

ii). Errors larger than 3xSD do occur, but with a less small frequency than might be expected. So whether the traditional error theory is valid in this area still needs to be proven.

iii). The mean of DH is usually not large, but it is not necessarily zero.

iv). The RMSE value of the DTM derived from the OS 1:63,360 scale topographic map will exceed $\pm$3.5m even in very flat areas. However, the range of RMSE values between $\pm$3.5 to 5.9m will still be quite acceptable for quite a number of small-scale applications.

As a final note about this study, it must be pointed out here that this experimental test was carried out early in 1987, shortly after the author started his Ph.D. study. At that time, it seemed still interesting to carry out such an experimental test. However, this study now seems a little out of date since the small scale DTM data base derived from the contours on the new OS 1:50,000 scale map series by the Mapping and Charting Establishment of Military Survey has already become available in some parts of England and will be available for the whole of Great Britain by 1991 as reported by Ley (1990) at the IBG Annual Conference held in the University of Glasgow earlier this year.

Nevertheless, these test results have been presented here since they still provide users with some useful information about the quality of the small scale DTM data which can be derived from the OS's smaller (1:50,000/63,360) scale maps. This may assist them in making a correct judgement as to whether such small scale DTM data is accurate enough for their specific applications. Undoubtedly, this data will still be of value to applications such as landscape visualisation, aircraft simulators, radio communications planning, etc. where large area coverage is required and a very high accuracy is not required of the DTM data.

13.3 Accuracy of the DTMs derived from photogrammetric contour data

The experimental test on the accuracy of the DTM derived from the OS 1:63,360 scale topographic maps has been reported in the previous section. In this section, an experimental test on the accuracy of the DTM derived from directly recorded photogrammetric contour data will be presented.

13.3.1 Test data

As has been stated in Chapter 9, a set of photogrammetrically measured contour data for each of the three test areas (Uppland, Sohnstetten and Spitze) was also made available to the author through the courtesy of Prof. H. Ebner and Dr. W. Reinhardt at the Technical University of Munich. These data sets were also measured on an analytical plotter. The contour intervals for each of these areas are given in Table 13.5, from which it will be seen that a contour interval (CI) of 5m was used for both the Uppland and Sohnstetten areas, while an interval of 1m was used for the Spitze area. A brief description of each of these areas has already been given in Table 9.1, while the **contour plots** and **isometric diagrams** for these areas have been given in Figs.9.1 to 9.3.

The data sets containing the information about **feature-specific points** and **feature-specific lines** have also been used in this test and the diagrammatic representations of these data sets have been given in Figs.9.4(b), 9.5(b) and 9.6(b).

The same **check points** as those used for the experimental tests which were presented in Chapter 9 have also been used in this test. Some useful information about these check points has already been given in Table 9.4, and the diagrammatic representations of the distributions of these check points have been given in Figs.9.7 to 9.9.

Table 13.5 Accuracy of the DTM data derived from ISPRS photogrammetrically measured contour data sets

	Uppland	Sohnstetten	Spitze
Contour Inter.	5.0 m	5.0 m	1.0 m
R M S E	1.74 m	0.91 m	0.27 m
Mean	1.05 m	0.22 m	0.10 m
S.D.	1.39 m	0.88 m	0.24 m
+ Emax	5.91 m	4.52 m	2.66 m
- Emax	-5.18 m	-3.01 m	-0.95 m

13.3.2 Test results using the contour data only

The PANACEA system was also used for this test and the same procedure as that used in the test of the OS contour data was applied to this ISPRS contour data. The accuracy results are shown in Table 13.5. Also as usual, the occurrence frequencies of the large residuals for each of the test areas have also been recorded and are given in Table 13.6.

Table 13.6 Occurrence frequencies of large residuals in the DTM derived from ISPRS photogrammetrically measured contour data

	Uppland		Sohnstetten		Spitze	
	No.	%	No.	%	No.	%
> 2 * SD	105	4.5	122	6.4	55	2.6
> 3 * SD	27	1.2	65	3.4	24	1.1
> 4 * SD	1	0.04	16	0.8	13	0.6

No regression analysis will be carried out for the accuracy results of the DTMs derived from photogrammetrically measured contour data since only one contour interval was used for each test area. Therefore, only a descriptive analysis of the accuracy figures and the residuals will be carried out.

From Table 13.5, it can be seen that the **RMSE or SD values** of the DTMs derived from the photogrammetrically measured contour data is more or less equal to CI/3. If these RMSE values are expressed in terms of per mil of flying height, then the RMSE values are 0.39 for the Uppland area, 0.61 for Sohnstetten and 0.45 for Spitze. These values are very much greater than 0.3 per mil of H which is the expected accuracy of dynamically measured data suggested by Rinner and Burkhardt (1972) and discussed in Chapter 5. Therefore, a big error budget appears to come from the loss in the fidelity of the terrain topography which is represented selectively by the measured contour lines. The RMSE values for such a budget are $\pm$0.25 per mil of H for the Uppland area, $\pm$0.53 for Sohnstetten area and $\pm$0.34 for Spitze area.

From the test results, it can be found that the **maximum errors** for the DTMs derived from photogrammetrically measured contour data are roughly equal to one contour interval. However, for the area of Spitze, there are 7 residuals which are larger than 1.0m which occur in a cluster. These errors occur in the area with terrain discontinuities already mentioned in Chapter 9 and therefore cannot be considered as being normal. After the deletion of these non-normal errors, the maximum

errors for this area are also about one contour interval. Errors greater than 4xSD did also occur, although their occurrence frequencies were small.

It is also interesting to relate the contour interval of a set of digital contour data to the grid interval of a set of measured gridded elevation data. The test results show that, for the Uppland area, the DTM data derived from the contour data with a 5m vertical interval cannot match the corresponding data derived from gridded spot height measurements even with the large grid interval of 80m. For the Spitze area, the contour data with 1m interval can only match the gridded elevation data with a grid interval of about 15m. However, with the same contour interval as that for Uppland, the contour data for Sohnstetten can match the grid data with a grid interval of about 30m. These results may be useful when making a decision as to which sampling strategy is to be used for a specific DTM project.

13.3.3 Test results using additional feature-specific data

However, if the data describing the **feature-specific points and lines** are added to the contour data, then the accuracy of the height points derived from the DTM can be greatly improved and the magnitude of the residual errors can also be significantly reduced. The accuracy results corresponding to such data sets are given in Table 13.7. Again, the occurrence frequencies of the large residuals have also been recorded and are given in Table 13.8.

Table 13.7 Accuracy of the DTM data derived from ISPRS photogrammetrically measured contour data plus feature-specific data

	Uppland	Sohnstetten	Spitze
Contour Inter.	5.0 m	5.0 m	1.0 m
R M S E	0.93 m	0.35 m	0.17 m
Mean	0.47 m	0.11 m	0.09 m
S.D.	0.80 m	0.35 m	0.15 m
+ Emax	3.25 m	1.73 m	0.75 m
- Emax	-5.18 m	-2.48 m	-0.95 m

It is interesting to note that, after the feature-specific data has been added, the accuracies of these DTMs derived from the ISPRS contour data have been greatly improved and the **values of RMSE or SD** have been reduced to a level of CI/5 from their original level of CI/3. The

actual figures for the improvement in the RMSE values for these three areas are about 45% for both the Uppland and Spitze areas and 60% for the Sohnstetten area. If the RMSE values of the DTMs obtained from the contour data plus the feature-specific information are expressed in terms of per mil of H, then the values are 0.21 for the Uppland area, 0.23 for Sohnstetten and 0.28 for Spitze. These values are even better than 0.3 per mil of H which is the expected value of photogrammetrically measured contour data discussed in Chapter 5.

Table 13.8 Occurrence frequncies of large residuals in DTM data derived from ISPRS photogrammetric contour data plus feature-specific data

	Uppland		Sohnstetten		Spitze	
	No.	%	No.	%	No.	%
> 2 * SD	129	5.6	96	5.1	71	3.4
> 3 * SD	21	0.9	20	1.1	16	0.8
> 4 * SD	3	0.1	7	0.4	7	0.3

The occurrence frequencies of **large residuals** are similar to those in the case when other photogrammetrically measured data sets were tested. When the feature-specific points and lines were added, then the magnitude of the large residuals can be greatly reduced and the DTM accuracy improved. However, errors exceeding 4xSD did still occur although the occurrence frequency is not large.

13.4 Concluding remarks

In map production, there are some specifications for the tolerable accuracy for the points interpolated from the contours on a map. Some of these specifications for topographic maps at different scales used in different countries have been given in Chapter 6. The accuracy figure for contours which is given by a map specification can, in fact, be considered as one kind of **expected accuracy** of the DTMs derived from a contour map because the DTM points are interpolated in a manner similar to that used for map checking.

Indeed, the results obtained from the two experimental tests presented in this chapter appear to prove this point. As one can imagine, the results obtained from these two tests are quite representative since each of the two main types of digital contour data (i.e. photogrammetrically measured contour data and digitised contour data from an older existing map series) were used for the tests carried out in this study, even though these tests were somewhat limited.

In the context of digital terrain modelling from digital contour data, one of the most important concerns is "how much **improvement** in the accuracy of the final DTM can be achieved if feature-specific points and lines are included ?". The results obtained from the tests using the ISPRS contour data in this study show that the RMSE or SD values can be reduced from a value of about CI/3 to a value of approximately CI/5. To put it in a relative sense, it might be said that the amount of improvement is about 40%. Indeed, this prediction may be considered as being a little conservative since the actual values obtained from the tests described in this chapter were in fact rather better than this value.

Actually, Tuladhar and Makarovic (1988) have also reported that an improvement of 53% in the RMSE value was achieved in their test when feature-specific points and lines were added to the digitised contour data in the generation of a DTM. This value is more or less similar to what has been obtained in the test described in this chapter.

Up to this stage, the problems of the accuracy of the DTMs derived from different data sources have all been touched upon in some way or another. Therefore, no more discussion about the DTM accuracy assessment will be carried out in this thesis. In the next chapter, a discussion concerning the optimum design of photogrammetric sampling will be carried out.

Chapter 14

Determination of Optimum Data Density with a Specified DTM Accuracy

Chapter Fourteen

Determination of Optimum Data Density with a Specified DTM Accuracy

14.1 Introduction

In the previous chapters, some investigations into the accuracy of DTMs have been made and a family of mathematical models have also been developed for the prediction of DTM accuracy. As has been pointed out in the introductory chapter (One), these mathematical models can be used as the basis to develop some procedures for the determination of the optimum data density used in a digital terrain modelling project. Indeed, this chapter is such an attempt to produce such procedures.

As has been discussed in Chapter 4, the density of DTM source data can be specified either in terms of sampling interval in the case of regularly distributed data or in terms of "number per unit area" in the case of irregularly distributed data. Therefore, in the former case, the minimum sampling interval defines the optimum data density and, in the latter, the optimum data density can be specified through the use of a minimum number of data points.

From a purely practical point of view, the following two topics which are related to optimum data density may be of interest to topographic scientists who are carrying out data acquisition or research work in the area of digital terrain modelling:-

i). Determination of the **optimum sampling interval** using an existing sampling strategy for a project with a specified accuracy requirement for the final DTM;

ii). Selection of a **minimum number** of data points from existing gridded data sets for a project with a specified accuracy requirement for the final DTM.

Actually, this chapter is an attempt to carry out some investigations into both of these two topics. For the first topic, only a theoretical discussion will be given since the experimental results obtained in Chapter 9 can be used as backup in this case. However, for the second topic, both a theoretical discussion and an experimental investigation will be carried out.

4.2 Determination of optimum sampling interval

First of all, the determination of the optimum sampling interval for both regular grid sampling and composite sampling will be discussed for

a project with a specified DTM accuracy requirement. Then, a discussion of the selection of a minimum number of data points from an existing data set will follow.

4.2.1 Introduction and background

The selection of an appropriate sampling interval is a very important concern for a DTM project since the use of too large a sampling interval will bring about the result that the required DTM accuracy cannot be fulfilled, while the use of too small a sampling interval will result in the collection of a huge amount of data which is not necessary for a specific DTM project. Therefore, the important thing is to find the sampling interval with which the given accuracy requirement can be reached and the minimum data excess will be generated. Such a sampling interval is referred to as **the optimum sampling interval** and the discussion of how to determine such an interval is the topic of this section.

Much time and energy has been spent on the determination of optimum sampling interval by several investigators, such as Ayeni (1982), Fritsch (1984, 1988), Balce (1986, 1987a, 1987b), Frederiksen et al (1986), Blais et al (1986), and others.

Balce (1986, 1987) used four programs, Spectra, Logkv, Linear, and RF, to test which is the most suitable interval for a practical project. **Spectra** is a program which carries out the test based on Fourier spectra analysis. **Logkv** is a program based on the operational use of variograms. The accuracy models on which these two programs have been based and developed have already been outlined in Chapter 12. **Linear** is a program based on a test of the reduction in accuracy which occurs if only the Nth points in a data set are selected. And **RF** is a program based on the analysis of mean slope or a roughness factor. Blais et al (1986) followed almost the same procedure. Both Balce and Blais et al found that no single method can be applied effectively to different types of terrain. And the values estimated by different programs differed greatly - by more than 2.5 times in some cases - which, in view of the discussions carried out in the previous chapters of this thesis, will come as no surprise.

Thus, an important question which must arise and be answered is why are the estimated values derived from these various models so different? The fact that the differences are so great means that at least some of them, if not all, must be unreliable. The problems which might have arisen with these programs could be:

i). They are designed for modelling only **profiles** but not a surface. However, as has been discussed previously (in Chapter 12), the

accuracy of the DTM points interpolated from profiles is quite different to those derived from a surface.

ii). The values used for **terrain roughness** are very sensitive to the lengths of profiles and to the variations in the elevations of the points measured along the profile themselves. This means that it is very difficult to find adequate or satisfactory values for terrain roughness for inclusion in these programs.

iii). The **six main factors** which affect the DTM accuracy (see Section 12.4) may not have been taken fully into consideration.

Actually, the determination of a value for the optimum sampling interval is always based on some kind of mathematical model of DTM accuracy into which is placed the required or specified accuracy for the DTM. However, most of the existing DTM accuracy models exhibit the various problems mentioned above. In this discussion, the accuracy models developed in Chapter 12 will be used instead. As has been pointed out there, these models are very simple, so only a relatively uncomplicated computation is needed. When the average slope of an area, the required DTM accuracy, and the accuracy with which source data can be measured are known, then the sampling interval can easily be calculated. The detailed discussion is given below.

14.2.2 Determination of sampling interval for composite sampling

As discussed in Chapter 12, the accuracy model for a composite data set is quite different to that for a gridded data set. Therefore, the optimum sampling intervals associated with a specified accuracy requirement are also differnet. In this section, the determination of an optimum sampling interval for composite sampling will be discussed while the determination of an optimum sampling interval for regular square grid sampling will be discussed in the next section.

It can also be found that there is not much difference between the accuracy model for use with bilinear surfaces and that for triangular facets, if both of them are being constructed from the same composite data set. Therefore, the accuracy models for the bilinear surfaces will be used here as an example. The mathematical expression can be recalled from Chapter 12 as follows:

$$VAR(DTM) = (4/9)\ VAR(PMD) + (5/768)\ (Dx\ \tan A)^2 \qquad (14.1)$$

Where VAR(DTM) is the specified DTM accuracy; VAR(PMD) is the accuracy of the photogrammetrically measured data (PMD); A is the average value of the slope angle for the area; and Dx is the sampling interval to be used. This equation can be rearranged as follows:

$$Dx = 12.39 \cot A \left(VAR(DTM) - 4/9\ VAR(PMD) \right)^{1/2} \qquad (14.2)$$

Since the value of Dx expressed by Equ.(14.2) is intended specifically for use with composite sampling, thus, a specific annotation, Dc, can be used. Then Equ.(14.2) can be rewritten as follows:

$$Dc = 12.39 \cot A \left(VAR(DTM) - 4/9\ VAR(PMD) \right)^{1/2} \qquad (14.3)$$

This formula can be used for planning a DTM project if **composite sampling** is being used. For example, for the Uppland area, if the DTM accuracy requirement is 0.75m, then it can be calculated that the optimum sampling interval, Dc, is about 70m.

14.2.3 Determination of sampling interval for square-grid sampling

It is assumed that there are no faults or terrain discontinuities inside the area to be modelled, since such features will be or should be measured in any case. The mathematical model for estimating the accuracy of DTM from gridded data sets is, as discussed in Chapter 12, as follows:

$$VAR(DTM) = (4/9)VAR(PMD) + (5/768)\left((1+4Dx/W)\ Dx \tan A \right)^2 \qquad (14.4)$$

Where, W is the so-called wavelength and a rough value of this parameter can be obtained from Equ.(12.27b). Equ.(14.4) can be rewritten as follows:

$$Dx\ (1 + 4Dx/W) = 12.39 \cot A \left(VAR(PMD) - (4/9)\ VAR(PMD) \right)^{1/2} = Dc \qquad (14.5)$$

Where Dc is the optimum sampling interval for composite sampling with a specified accuracy requirement of VAR(PMD) (see Equ.(14.3)). Using Dr to denote the optimum sampling for regular grid sampling, then the exact solution to Equ.(14.5) is as follows:

$$Dr = \left((1 + 16\ Dc/W \right)^{1/2} - 1) \ /\ (8/W) \qquad (14.6)$$

It seems that this expression is a little inconvenient to use from the computational point of view. In practice, a more convenient expression which will result in an **approximate value** can be obtained from Equ.(14.5). Such a value is derived in the following paragraphs.

First of all, Equ.(14.5) can be rewritten in a form as follows:

$$Dx = Dc \ /\ (1 + 4\ Dx\ /\ W) \qquad (14.7)$$

At first sight, it would appear that this expression is not convenient at all since the variable to be solved - Dx - appears on both sides of the equation. However, it is not the final expression. The aim of this discussion is to derive a formula which is similar to Equ.(14.7) and is capable of producing a value which is very close to that produced by Equ.(14.6) for Dx.

From Equ.(14.7), it can be seen clearly that the value of the optimum sampling interval for regular grid sampling is only a fraction of that for composite sampling. If an approximate value, say Da, is used to replace the Dx in the right side of Equ.(14.7), then an approximate value can be obtained for the optimum sampling interval. Therefore, such an approximate formula can be expressed as follows:

$$Dr = Dc / (1 + 4\ Da / W) \qquad (14.8)$$

Actually, at the initial stage, the value of Dc may be used for this approximate value, Da. Taking the example used in the previous section, it was computed that Dc=70m. By letting Da=70m, it can be computed from Equ.(14.8) that Dr=44m. The exact solution obtained from Equ.(14.6) is that Dr=49.3m. Therefore, the difference between the approximate solution and exact solution is about 5m. However, if one likes, a **solution closer to the exact solution** can be obtained from Equ.(14.8) by substituting the the variable Da with the new value of Dr which has just been computed from Equ.(14.8), 44m in this particular example. Thus, a new value of Dr=50m can be obtained from Equ.(14.8). Actually, if such an iteration is repeated, then the approximate value will approach to the exact one. For example, the values of Dr in the next three iterations will be equal to 49.1, 49.4 and 49.3, respectively.

14.3 Selection of a minimum number of data points

In the previous section, the matter regarding the determination of optimum sampling interval for both square grid sampling and composite sampling has been discussed. In this section, attention is turned to the use of a minimum number of data points which are irregularly located for the modelling of a DTM surface with a specified accuracy requirement. The main concern is with the selection of such an irregularly distributed data set from a very dense set of gridded data.

14.3.1 Introduction and background

As has been discussed in Chapter 8, in the case of data acquisition using **digital image correlation** techniques, the resulting data sets will be very dense, for example, up to 500,000 to 700,000 points per stereo-model are measured in the case of the GPM-2 (Petrie, 1990). Such

very dense data sets are not always suitable or appropriate for use by applications specialists such as engineers, planners, etc; indeed in some cases, their sheer volume constitutes a definite deterrent or drawback to their use. Also if the efficiency of the modelling process and the cost of computation are being considered, as they should be, then a filtering procedure needs to be applied to such data sets so that only a minimum number of data points will be selected from the data set while the specified accuracy of the final DTM will still be achieved.

A filtering processing may be also applicable in other cases when **regular grid sampling** has been employed as the sampling strategy for data acquisition. In this case, the usual practice is to employ a sampling interval which is suitable for the area with the roughest terrain in order to ensure the accuracy of the final DTM. Of course, such an interval can be determined by using the precedures described in the previous section. However, the problem is that, in the flatter areas, too many data points will have been measured. Therefore, it may be also desirable to apply a filtering processing to such a data set.

One may argue that **progressive sampling** can be used to avoid such a problem. Indeed, the idea of using varying grid cell sizes for each piece of terrain with a different roughness is that employed in the progressive sampling strategy originally suggested by Makarovic (1973). This sampling strategy has been implemented by many organisations and works well in many cases. However, progressive sampling cannot solve the whole problem.

At the DTM workshop held at Edmonton, Canada in 1984 (see Toomey, 1984), there were a lot of interesting discussions about this sampling strategy based on the experinces gained by the participants. Prof. Collins commented that "the **fundamental problem** of progressive sampling is that a low second derivative value does not mean that the surface is necessarily OK. When a new point is tested, it may be good enough; but that doesn't mean that there are no irregularities in the neighbourhood." Prof. Molnar agreed with Collins and stated that "the remarks of Professor Collins are really important. You can check whether it is OK but the answer doesn't give you a 100% security." These remarks indicate that it might be difficult to come to a reliable decision as to whether or not an area should be further densified. Indeed alternative criteria for making such a decision have been searched for by Makarovic himself (1975) and by one of his students (Charif and Makarovic, 1988). However, this foundamental problem is one which is difficult to solve since this methodology goes from the unknown to the known. Therefore, it may be concluded that simple regular grid sampling will still remain in widespread use in DTM practice and that the problems associated with this sampling strategy still need to be

considered and an alternative solution offered.

From the discussions conducted above, it does seem that an investigation into the selection of a minimum number of points from data sets such as those acquired from digital image correlation and dense regular grid sampling is very desirable. Thus an effort to contribute a solution to this problem has been carried out in this project.

14.3.2 The procedure used for data selection

As has been discussed in Chapter 4, a point which has been specially selected on the basis of its importance for terrain representation has a greater importance or significance than a purely random point in terms of defining or representing a particular surface area. In order to select such points, the significance of every point needs first of all to be evaluated.

Chen and Guevara (1987) developed a procedure, which defines Very Important Points (VIPs), to serve such a purpose. The first step in their procedure is to select a measure for significance. The measure which was used by them is the **spatial differential operator**, which is a measure of the changing behaviour of a point from its neighbours. The principle for the one-dimension case is as follows:

Suppose, the height of a point along a profile is a function of its position as follows:-

$$H = f(x) \tag{14.9}$$

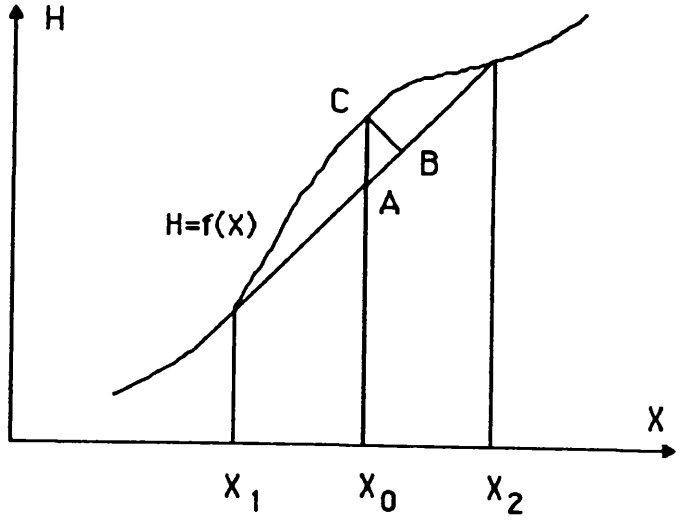


Fig.14.1 The second differential of a point

then, its second order differential value at point X_0 (Fig.14.1) is

$$d^2H/dX^2 = f''(X_0) \qquad (14.10)$$
$$= 2\ (\ f(X_0) - (f(X_1) + f(X_2))/2\ )$$

Actually, the distance AC in Fig.14.1 is the second order differential value at point X_0. They also consider four spatial directions (up-down, left-right, upper left-lower right, lower left-upper right). For every point, the second differential values for all four directions are added together to represent the **degree of the significance** of this point. Actually, the line of thought is very similar to that used by Makarovic (1973) in designing the progressive sampling method.

Chen and Guevara also pointed out that, if the number of points to be selected is specified, then those points with the greatest significance can be selected. However, in the case of this present study, what has been defined and is therefore known is the required DTM accuracy. Therefore, the selection of VIPs should be related to the required accuracy of the DTM itself instead of a pre-defined number of points. In this case, a **critical value** should be used so that any point with a degree of significance smaller than this value may be removed.

14.3.3 The relationship between the critical value used for data selection and the resulting DTM accuracy loss

To find the **critical value** for data selection, a close examination of distance AC in Fig.14.1 must first be undertaken. It can be found that AC is the error at $X=X_0$ if X_0 is removed and the profile is constructed by linear interpolation between the elevation values at X_1 and X_2. This means that the value of the so-called degree of significance itself represents the DTM error and a loss of accuracy will result from the selection of VIPs or the removal of those points which have been regarded as insignificant. The problem now arising is how much will the loss of accuracy be in terms of standard deviation or RMSE if all the data points with the so-called degree of significance smaller than a specific value are removed. In other words, the **relationship** between the accuracy loss and the specified critical value needs to be investigated.

If the distribution of these errors is known, then such a relationship can easily be set out. The distribution may be known by analysing these errors. Obviously, their distribution may vary from area to area. Therefore, some kind of rough estimation has been carried out in this study.

To find such a relationship, an experimental investigation has been carried out in this study. The results of these tests will be reported in the next section. By analysing the set of values for the degree of

significance for the data points obtained from the experimental tests and using a trial-and-error method, it has been found that the accuracy loss due to the deletion of those insignificant points in terms of standard deviation is 1/3 of the critical (threshold) value used.

Suppose that the critical value of the degree of significance for data selection is SIG(crit), while SD(loss) is used to denote the accuracy loss due to the deletion of insignificant points in terms of standard deviation, then the following expression can be obtained:-

$$SD(loss) = SIG(crit) / 3 \qquad (14.11)$$

In practice, since the tolerable value of such an accuracy loss may be known, the critical value, SIG(crit), for data selection can then be determined using Equ.(14.11).

14.3.4 Experimental Results

In order to find the relationship between the critical value used for data selection and the resulting DTM accuracy loss, some experimental tests have been carried out. Two of the data sets described in Chapter 9 were used, i.e. the sets of gridded data for both the Uppland and Sohnstetten areas.

In this test, the value of the so-called degree of significance of every point is first computed; then different critical values are specified for data selection. The accuracy of the DTM derived from the data sets after the removal of the insignificant (i.e. unwanted) data points is then computed to see how the experimental results fit the theoretically expected values.

Table 14.1 lists the results of DTM accuracy obtained from the data sets selected from one of the two grids for the Uppland area with different critical values. In this table, "Pts No." denotes the number of gridded data points which were selected; "% No of Pts" denotes the percentage value of grid points selected from the original total number of 1,862; and, as usual, RMSE, SD and Mean denote the root mean square error, standard deviation and the mean of the errors in the DTM from these selected grid points. Under the term "Orig", all the values used are either those of the original data set or were obtained from the original data sets. For example, it can be found that the root mean square error and standard deviation values obtained from the original data set are $\pm$0.770m and $\pm$0.764m, respectively.

This table lists the results of the accuracy of the DTMs derived from selected data sets with the threshold values from 0.4m to 0.9m. The results in this table demonstrate clearly that, with an increase in the

value of SIG(crit), the number of data points selected by this filtering procedure is fewer; thus the accuracy of the DTM derived from the selected data set becomes lower (i.e. the values for RMSE and SD are greater).

Table 14.1 DTM accuracy variation with difference requirements for grid 1 of the Uppland area

Data Description				Accuracy Description		
File	SIG(crit)	Pts No.	% No of Pts	RMSE	SD	Mean
Orig	0.0m	1,862	100.0%	0.770	0.764	0.102
F4	0.4m	1,489	80.0%	0.778	0.772	0.098
F5	0.5m	1,392	74.8%	0.779	0.775	0.086
F6	0.6m	1,301	69.9%	0.783	0.787	0.084
F7	0.7m	1,206	64.8%	0.801	0.795	0.102
F8	0.8m	1,123	60.3%	0.810	0.803	0.104
F9	0.9m	1,044	56.1%	0.825	0.818	0.108

A comparison of the experimental results (only SDs) with the values predicted computed from Equ.(14.11) is shown in Table 14.2. From this table, it can be seen clearly that Equ.(14.11) is almost a perfect expression of the results obtained from this test.

Table 14.2 DTM precision variation with critical value for the Uppland area

SIG(crit)	0.4m	0.5m	0.6m	0.7m	0.8m	0.9m
Experimental	0.772m	0.777m	0.787	0.795m	0.803m	0.818m
Theoretical	0.776m	0.782m	0.790m	0.799m	0.809m	0.821m

In order to come to a more reliable conclusion, a second test was also carried out using the data for the Sohnstetten area. The test results are listed in Table 14.3. For the original data set, the RMSE and SD values are $\pm$0.572m and $\pm$0.561m, respectively. The range of threshold values used in this test range from 0.2m to 0.7m. The comparison of the experimental results with the values predicted by Equ.(14.11) is shown in Table 14.4. Again, only the SD values are listed, and as Table 14.4 shows, it was found that the values predicted by Equ.(14.11) agree almost with the experimental test results.

Table 14.3 DTM accuracy variation with difference requirements for grid data of the Sohnstetten area

Data Description				Accuracy Description		
File	SIG(crit)	Pts No.	% No of Pts	RMSE	SD	Mean
Orig	0.0m	1,716	100.0%	0.572	0.561	-0.112
F2	0.2m	1,442	84.0%	0.579	0.567	-0.114
F3	0.3m	1,321	77.0%	0.581	0.571	-0.112
F4	0.4m	1,220	71.1%	0.580	0.571	-0.102
F5	0.5m	1,110	64.7%	0.586	0.576	-0.106
F6	0.6m	1,017	59.3%	0.596	0.588	-0.095
F7	0.7m	942	54.9%	0.613	0.606	-0.087

Tab.14.4 DTM precision variation with critical value for the Sohnstetten area

SIG(crit)	0.2m	0.3m	0.4m	0.5m	0.6m	0.7m
Experimental	0.567m	0.571m	0.571m	0.576m	0.588m	0.606m
Theoretical	0.565m	0.570m	0.577m	0.585m	0.596m	0.608m

14.3.5 Summary and discussion

In this section, a filtering procedure has been described for the selection of a minimum number of data points from a dense gridded data set using the second differential value as the measure of the degree of significance. It has also been shown that the important problem with such a filtering procedure is to decide on a critical value for data selection.

It has also been pointed out that there will be an accuracy loss in the final DTM resulting from the removal of those points which have been regarded as being insignificant. Therefore, the determination of the critical value should be related to the tolerable value of such an accuracy loss.

In a relative sense, the importance of a data point in surface representation will depend on the closeness of the data points. If all the data points are located very closely together, then the removal of an individual data point can only result in an insignificant loss of accuracy in the final DTM. By contrast, the removal of an individual data point might bring about a significant accuracy loss when the interval between points is very large. Actually, the measure of the degree of significance used in this study has already taken this fact into consideration.

Of course, in a manner similar to the problem associated with progressive sampling discussed above in Section 14.3.1, there may still be a problem of reliability in the decision as to which point should be removed. However, the decision made in this case will definitely be more reliable since this method goes from a dense data set to a coarse data set.

14.4 Optimization of regular grid sampling with compressive approach

As has been discussed in the previous section, there will be always a problem associated with regular grid sampling in that the data points measured in the flatter areas will be too dense. Of course, a data filtering procedure such as the one which has just been described in the previous section can be applied to this data. Moreover, the matter of the accuracy loss in the final DTM when those points which have been evaluated as being insignificant are removed from the data set has also been investigated. The relationship between the critical value used for data selection and the resulting accuracy loss in the final DTM has been expressed by Equ.(14.11). Therefore, in the case of regular grid sampling, if a decision has been made to apply such a filtering procedure to the acquired data set, then the extent of the accuracy loss in the final DTM should also be considered at the time when the optimum interval for the grid sampling is being determined.

Suppose that the required final DTM accuracy in terms of variance is VAR(DTM), and the accuracy loss due to the deletion of the measured points which have been considered as being insignificant is VAR(loss), Then, the accuracy requirement for the initial regular grid is:-

$$VAR(inital) = VAR(DTM) - VAR(loss) \qquad (14.12)$$

The procedures described earlier in this chapter can be used for such a purpose. In this way, an overall optimization of regular grid sampling might be reached.

14.5 Concluding remarks

In this chapter, the determination of an optimum sampling interval for both regular grid sampling and composite sampling has first of all been discussed; then, the problem of selection of data points from a gridded data set with an optimum data density has been investigated; and finally an integration of regular grid sampling and a data filtering technique has been outlined.

Up to this stage, all the main goals which have been set out in the introductory chapter (One) have been reached at least in a certain fashion. Thus, the time has now been reached to make some concluding remarks on the investigations which have been carried out in this project. These remarks will be presented in the next chapter.

Chapter 15

Concluding Remarks

Chapter Fifteen

Concluding Remarks

The investigations carried out in this project have all been documented in the previous chapters. At this stage, it seems pertinent to have some concluding remarks on these investigations to see how much contribution has been made to the theory and practice of digital terrain modelling; how adequate are the results obtained from the various investigations; and what kind of limitations exist in the present investigations. After such a discussion, some recommendations might be made for future research.

15.1 General remarks on the research carried out in this project

It has been pointed out in the introductory chapter (One) that there are **three main concerns** with digital terrain modelling, i.e. accuracy, cost and efficiency. Among them, **accuracy** is usually the core. Therefore, the accuracy of the final DTM surface has been the main matter which has been pursued in this project. It has also been pointed out previously that the accuracy of the final DTM surface is affected by many factors such as the characteristics of the terrain surface, the sampling pattern used and the surface reconstruction. It has also been found by previous investigators that **sampling** is the most vital factor. Therefore, sampling has also been an important issue which has been addressed in this project. Thus, this thesis is entitled "sampling strategy and accuracy assessment for digital terrain modelling".

The **accuracy** of the surfaces produced by a DTM is a complex matter. In order to clarify this matter, first of all, the factors which affect the accuracy of the DTM surfaces, need to be clearly defined and examined. The discussions in Chapters 3, 4, 5, 6 and 7, which form the first part of this thesis, are an attempt to serve such a purpose. In other words, the discussions carried out in these chapters laid down a solid **theoretical foundation** for the later discussions of the assessment of DTM accuracy.

The main goal arising from the theoretical discussions and the accompanying experimental work on the assessment of DTM accuracy in this project is to develop a family of **mathematical models** which will define or predict the accuracy of the different types of digital terrain model surface which are constructed from different data patterns. This goal may be achieved in two ways, i.e. either by a theoretical analysis or by experimental tests, or a combination of both. The purely experimental method is time consuming and costly, in

which case, the method using a theoretical analysis is more attractive. Therefore, the mathematical models in this project were developed largely through a **theoretical analysis** (Chapter 12).

To attempt to establish mathematical models for DTM accuracy, some discussions of the general concept of a mathematical model and the theoretical background to the development of mathematical models are also required as a guideline. Such discussions have been conducted in Chapter 2. To validate a mathematical model, some practical results obtained from **experimental tests** are required. Therefore, a specially designed experimental investigation into the variation of the DTM accuracy with sampling interval was also carried out in this project and reported in Chapter 9. Also an experimental test on the accuracy of a DTM acquired from space photography taken by the Metric Camera has been carried out and reported in Chapter 10. In addition, a discussion about the effects of the check points used in an experimental test of DTM accuracy on the reliability of the estimated DTM accuracy figures which were obtained from the experimental test has also been carried out in Chapter 11.

The investigations mentioned above were all concerned with the accuracy of the DTMs derived from photogrammetrically measured data. However, some other experimental investigations into the accuracy of the DTMs derived from digital contour data have also been carried out and the results have been reported in Chapter 13.

As has been pointed out previously, the **errors in the source data** affect the accuracy of the final DTM. In the case of DTM accuracy assessment through a theoretical analysis, it is assumed implicitly that the source data which is to be used for modelling the terrain surface is free from gross errors. However, in practice, the measured raw (source) data - especially that derived from digital image correlation techniques - will include gross errors. It goes virtually without saying that these **gross errors** should be detected and removed from the source data set. In the studies carried out in this project, some numerical algorithms were developed to perform such an operation. Furthermore, an investigation into the effect of **random errors** present in the source data on the DTM quality has also been carried out and also reported in Chapter 8.

The establishment of a family of mathematical models for DTM accuracy is not the final destination of this project, but instead it is only a means to reach this destination. Indeed, the **final destination** of this project is the optimisation of photogrammetric sampling for a DTM project with a given accuracy requirement, i.e. to optimize the sampling strategy and to determine the optimum sampling interval for a particular sampling strategy.

Optimisation of a sampling strategy based on the use of photogrammetric methods is also a difficult task. In order that an optimized sampling strategy could be devised, a deep insight into the matter of photogrammetric sampling is essential. Chapter 4 attempts to provide this. With the theoretical background to photogrammetric sampling given in Chapter 4 and the DTM accuracy models described in Chapter 12, the vexed problem of the optimisation of photogrammetric sampling could then be discussed. Such a discussion has been carried out at the end of this thesis (in Chapter 14).

15.2 Remarks on mathematical models of DTM accuracy

In the context of DTM accuracy, for a time in the late 1970s and early 1980s, there was a **trend** that more and more complicated mathematical models should be devised and used. However, as has been pointed out previously in Chapter 2, a complicated mathematical model is not always necessary although the phenomena under investigation could be complicated. Therefore, this trend towards compication seemed to suggest that the problem as to how the errors from different sources are propagated in the processes of digital terrain modelling was (and is) still not fully understood. As has also been pointed out in Chapter 2, this trend indicates that, if possible, an entirely different family of mathematical models needs to be considered. Hopefully, this explains why the author has concentrated much of his effort into development of new mathematical models.

Each of the mathematical models of DTM accuracy developed in this project consists of **two terms**, i.e. a term related to the accuracy of source data and a term related to the sampling interval and terrain characteristics. As has been mentioned previously, the mathematical models of DTM accuracy developed in this project are all simple, accurate, precise, descriptively realistic, robust, and fruitful, but **not general**. More concretely, they should not be applied when a very large sampling interval is used. Therefore, further research work is needed to extend the applicability of this family of accuracy models to such a case (with a large sampling interval).

As one can imagine, in the case of a **large sampling interval**, the rate at which the value of the standard deviation of the final DTM increases with the increase in sampling interval will become lower and lower after the sampling interval reaches a certain level. In topographic mapping by ground surveying, some test results (Li, 1982) have come to suggest that the loss in the accuracy of contour lines due to the effect of topographic generalisation (the loss in fidelity of the terrain topography which is represented by the data points measured by

ground survey) can be approximated using the following expression:

$$SD(c) = k. L^{1/2} \qquad \text{(unit: m)} \qquad (15.1)$$

Where, k is a constant and L is the average point interval in units of 10m. Such an expression might also be suitable for DTM points in the case when a large sampling interval is used. Actually, a trend similar to that expressed by Equ.(15.1) has also been illustrated by Ley (1990) in the IBG Annual Conference held at the University of Glasgow earlier this year (January 1990). Nevertheless, more investigations into this matter are still required to clarify this point.

Regarding the **assumption** about a relationship between the extreme values of a distribution and the corresponding standard deviation, on the basis of which a family of mathematical models of DTM accuracy have been developed in Chapter 12, a comparison with the the limited test results described in Chapter 9 has shown that it is descriptively realistic. However, many more investigations of this type are desirable since such a relationship may vary with the characteristics of the terrain surface. In such a study, the terrain characteristics may then be classified according to their geomorphological forms and an investigation made into such a relationship for each individual type of terrain. In this way, more accurate models may be obtained.

It has also been mentioned previously in Chapter 4, that the **distribution** of source data will also affect the DTM accuracy. In fact, the distribution of the raw (source) data points can be defined by two parameters, i.e. their location and pattern. In the case of regular grid sampling and composite sampling, the location of source data may be defined by the **orientation** of the grids. In the study carried out in this project, only pattern has been taken into serious consideration, since it was assumed that the orientation is not among the most important factors affecting DTM accuracy. Nevertheless, it must have some kind of effect on the DTM accuracy even though the effect might not be too significant. Therefore, a family of more comprehensive mathematical models may be developed by taking the **location** of the raw (source) data points into consideration. Also some investigations into the DTM accuracy related to **other patterns** of DTM source data may be of interest.

The mathematical models of DTM accuracy developed in this project are intended for use with the DTMs derived from photogrammetrically measured data, but not for the DTMs derived from **digital contour data.** For the accuracy of the DTMs derived from contour data, it has been found from the test results described in Chapter 13 that the map specifications can be used as empirical models. This empirical model consists of two parts, i.e. a value of CI/K plus a term related to the

average slope angle of the area. From the limited tests carried out in this project, it has been found that the coefficient K is approximately equal to 3 for the DTMs derived from digital contour data only. As expected, the accuracy of the DTMs derived from the digital contour data will be greatly improved if the feature-specific points and lines (e.g. derived from photogrammetric measurement) are also included. The limited test results show that the value of the coefficient K will be increased to 5. However, as one can imagine, the amount of improvement in the accuracy of the final DTM will also vary with the characteristics of the terrain surface itself and obviously the value of 5 for the K is only an approximate value. Therefore, more investigations into the amount of such an improvement for different types of terrain would be of interest to establish a suitable range of values.

When a test of the accuracy of the DTMs derived from digital contour data is carried out, usually, the source data (i.e. digital contour data) is assumed to be so dense that it can represent the original terrain sufficiently well. In this way, a huge amount of digital data may result. However, sometimes, not all of the digital contour data points are necessary for a given degree of DTM accuracy. Thus there may be a redundancy of data, which needs to be removed using some methods such as those reviewed and the one developed by the author (Li, 1988b). An interesting point arising here is "how does the accuracy of the DTMs vary with the **density** of the retained digital contour data while different methods are employed to reduce the data density"? Actually, Shearer and one of his students (Gar-Al-Nabi, 1988) have already carried out some investigations into the accuracy of the DTM varying with the interval between the pairs of digital contour data points. Shearer has given an interesting demonstration of their results in the IBG Annual Conference held at the University of Glasgow earlier this year (Shearer and Li, 1990). However, more investigations into this area are still desirable.

As an extension of this project, the mathematical models of the **accuracy of DTM products** such as contour maps may also be considered here. Ostman (1987) tried to form some mathematical models for the accuracy of DTM products using different terrain descriptors such as correlation functions, autocovariance functions and others. However, as has already been pointed out in Chapter 3, these descriptors have many demerits. Therefore, the "goodness" of these models still need to be evaluated. On the other hand, a family of mathematical models for the accuracy of DTM products may also be developed in a way similar to the theoretical analysis carried out in this project.

15.3 Remarks on photogrammetric sampling strategy

Since more and more evidences is becoming available to show that sampling is the most vital stage in digital terrain modelling, therefore, a comprehensive sampling strategy together with the establishment or definition of an optimum sampling interval is of most importance for any DTM project. And this, as has already been pointed out previously at start of this thesis, is the final goal of this project.

Regarding a **sampling strategy**, there are two important aspects to be investigated. The first is how to optimize existing sampling strategies and the second is to design a comprehensive new sampling strategy.

The **optimization** of a sampling strategy means that it has to satisfy the requirements discussed in Chapter 4. Actually, Charif and Makarovic (1988) have already reported a relevant investigation. They tried to establish some rules and efficient procedures for the selection of feature-specific points and feature-specific lines in composite sampling and to study the effects of different decision models for the densification of grids in progressive sampling. However, only limited results have been published as yet since their investigations were still in progress at the time when they produced their report. Therefore, some further investigations into the optimization of different (existing) sampling strategies would undoubtedly be of value.

The idea of using varying sizes of grid cell for terrain with different roughness employed in **progressive sampling** suggested by Makarovic (1973) is potentially very beneficial. However, as has been discussed previously in Chapter 14, this sampling strategy still cannot solve the whole problem of defining optimum data density. Therefore, it would also be of value to seek an alternative solution.

By borrowing the idea of measuring the DTM source data with varying densities for different areas with different roughness of the terrain topography, an alternative to progressive sampling - going from a fine mesh to coarse one - may be used. Actually, the **compressive approach** comprising two stages of sampling discussed in Chapter 14 belongs to this alternative. Of course, for the post-measurement filtering, various procedures based on different criteria may be used. For example, the comprehensive curvature scheme used by Dikau (see Chapter 3) could well be an alternative to the second derivative which has been used in this project. Therefore, an intensive investigation into the possible alternatives for compressing or reducing the amount of data which needs to be measured might also be an interesting topic.

For the time being, in the discussions carried out by various investigators including the present author, only the accuracy of the final DTMs has been taken into consideration in the design of a photogrammetric sampling strategy. However, two other factors, i.e. **efficiency and cost**, should also be taken into account. Only in this way, can an overall optimization of sampling strategy be reached. For the modelling of a small area, this might not appear to be very important. However, when a national DTM data bank is considered, this overall optimization is indeed a very important concern. In this respect, the experiences gained by those people who have been involved in the production of DTMs at a regional or provincial scale will be extremely important. As an example, the experience gained in the modelling of Alberta province (see Toomey, 1988) and that being gained by the Mapping and Charting Establishment here in the UK (Ley, 1990) will be of great value.

15.4 Final remarks

In the previous sections, some discussions related to this project have been conducted. Indeed, some of them may be regarded as comments; some as notices; some as observations; while others may be considered as being merely the author's thoughts or opinions. However, no matter what they really belong to, the English word "remarks" covers all these meanings (see the Collins English Dictionary, p719). Therefore, this chapter is entitled "concluding remarks".

Up to this stage, the author has already said what he would like to say about his research project. As a final remark, the author would like to apologise for the non-standard mathematical annotations that have been used throughout this thesis, which may be an inconvenience in reading, and for the possible mistakes which have inevitably been left in this thesis. The author also hopes that some useful information may be obtained from this thesis.

BIBLIOGRAPHY

BIBLIOGRAPHY

Ackermann, F., 1979. The accuracy of digital terrain models. **Proceedings of 37th Photogrammetric Week**, University of Stuttgart: 113-143.

Ackermann, F. & Stark, E., 1985. Digital elevation model from Spacelab Metric Camera photographs. **Metric Camera Workshop**, 9-12.

Alspaugh, D., 1985. **Analysis of Dynamic Terrain Profile on Analytical Photogrammetric Systems.** Ph.D. Thesis, Purdue University. 243pp.

Ayeni, O., 1976. Objective terrain description and classification for digital terrain model. **International Archives of Photogrammetry and Remote Sensing**, 23(III/1).

Ayeni, O., 1982. Optimal sampling for digital terrain models: A trend towards automation. **Photogrammetric Engineering and Remote Sensing**, 48(11):1687-1694.

Azizi, A., 1990. **The Design and Implementation of a Digital Stereo-Photogrammetric System on the IBM 3090 Multi-User Machine.** Ph.D. Thesis, University of Glasgow. 235pp.

Baffisfore, A., 1957. A do-it-yourself terrain model. **Photogrammetric Engineering**, 23:712-720.

Balce, A., 1986. Determination of optimum sampling interval in grid sampling of DTM for large-scale application. **International Archives of Photogrammetry and Remote Sensing**, 26(3):40-53.

Balce, A., 1987a. Determination of optimum sampling interval in grid digital elevation models (DEM) data acquisition. **Photogrammetric Engineering and Remote Sensing**, 53(3):323-330.

Balce, A., 1987b. Quality control of height accuracy of digital elevation models. **ITC Journal**, 1987-4:327-332.

Beerenwinkel, R., Bonjour, R., Hersch, R. & Kolbl, O., 1986. Real-time stereo image injection for photogrammetric plotting. **International Archives of Photogrammetry and Remote Sensing,** 26(4):99-109.

Berg, G., 1988. Investigations on the measurement accuracy of the analytical plotter Wild Aviolyt AC-1. **International Archives of Photogrammetry and Remote Sensing,** 27(B2):35-43.

Blais, J., Chapman, M. & Lam, W., 1986. Optimum sampling interval in theory and practice. **Proceedings of 2nd International Symposium on Spatial Data Handling,** 185-192.

Brassel, K. & Reif,D., 1979. Procedure to generate Thiessen polygon. **Geographical Analysis,** 11(3):289-303.

Burington, R. & May,D., 1970. **Handbook of Probability and Statistics with Tables.** 2nd edition. McGraw-Hill Book Company. 462pp.

Charif, M. & Makarovic,B., 1988. Optimizing progressive and composite sampling for digital terrain model. **International Archives of Photogrammetry and Remote Sensing,** 27(B10):III-264-280.
Also **ITC Journal,** 1989-2:104-111.

Chen, Z. & Guevara,J., 1987. Systematic selection of very important points (VIP) from digital terrain model for constructing triangular irregular networks. **Auto Carto 8,** 50-56.

Chestnut, H., 1965. **System Engineering Tools.** John Wiley & Sons, Inc. N.Y., London, Sydney. 162pp.

Chorley, R, 1987. **Handling Geographic Information.** Report of the Committee of Enquiry Chaired by Lord Chorley. 208pp.

Christensen, A., 1987. Fitting a triangulation to contour lines. **Auto Carto 8,** 57-67.

Clarke, A., Gruen, A. & Loon, J., 1982. A contour-specific interpolation algorithm for DEM generation. **International Archives of Photogrammetry and Remote Sensing,** 24(III):68-81.

Cole, G., Dixon, R. & Sorrel, P., 1983. **The First Step in Cartography.** Working Paper No.7, North East London Polytechnic. 126pp.

Cryer, J., 1986. **Time Series Analysis with Minitab.** Duxbury Press. 286pp.

David, M., 1977. **Geostatistical Ore Reserve Estimation.** Elsevier Scientific Publishing Company. 364pp.

Delaunay, B., 1934., Sur la sphere vide. **Bulletin of the Academy of Sciences of the USSR,** Classe. Sci. Math. Nat: 793-800.

Demek, J., 1972. **Manual of Detailed Geomorphological Mapping.** Academia, Prague. 344pp.

Dikau, R., 1989. **The Application of a Digital Relief Model to Landform Analysis in Geomorphology.** in "Three Dimensional Applications in Geographic Information Systems", edited by J. Raper. Taylor & Francis. 51-78.

Dummer, G. & Winton, R., 1986. **An Elementary Guide to Reliability.** 3rd edition. Pergamon Press. 47pp.

Ebisch, K., 1984. Effect of digital elevation model resolution on the properties of contours. **Technical Papers, ASP-ACSM Fall Convention,** 424-434.

Ebner, H., Hofmann-Wellenhof, B., Reiss, P. & Steidler, F., 1980. HIFI - A minicomputer program package for height interpolation by finite elements. **International Archives of Photogrammetry and Remote Sensing,** 23(IV):202-241.

Ebner, H. & Reinhardt,W., 1987. Verification of DTM data acquisition and digital contours by means of optical superimposition. **Technical Papers, ASPRS-ACSM Fall Convention,** 2:82-88.

Elfick, M., 1979. Contouring by use of a triangular mesh. **The Cartographic Journal**, 16:24-29.

Elmhorst, A. & Muller, W., 1988. Generation of DTMs with space photographs. **International Archives of Photogrammetry and Remote Sensing**, 27(B10):II-352-361.

El-Niweiri, A., 1988. **Geometric Accuracy Testing, Evaluation and Applicability of Space Imagery to the Small Scale Topographic Mapping of the Sudan.** Ph.D. Thesis. University of Glasgow. 380pp.

Emshoff, J. & Sisson, R., 1970. **Design and Use of Computer Simulation Models.** The Macmillan Company, London. 302pp.

Evans, I., 1972. **General Geomorphometry, Derivatives of Altitude, and the Descriptive Statistics**, in "Spatial Analysis in Geomorphology" edited by R. Chorley. Methuen & Co. Ltd., London. 17-90.

Evans, I., 1981. **General Geomorphometry**, in "Geomorphlogical Techniques" edited by A. Goudie. George Allen & Unwin. Boston and Sydney. 31-37.

Fagan, P., 1972. Photogrammetry in the National Survey. **Photogrammetric Record**, 7(40):405-427.

Frederiksen, P., 1980. Terrain analysis and accuracy prediction by means of the Fourier transformation. **International Archives of Photogrammetry and Remote Sensing**, 23(4):284-293.
Also **Photogrammetria**, 36(1981):145-157.

Frederiksen, P., Jacobi, O. & Justesen, J., 1978. Fourier transformation von hohenbeobachtungen. **ZFV**, 103:64-79.

Frederiksen, P., Jacobi, O. & Kubik, K., 1983. Measuring terrain roughness by topographic dimension. **Proceedings of International Colloquium on Mathematical Aspects of DEM.** Stockholm.

Frederiksen, P., Jacobi, O. & Kubik, K., 1985. A review of current trends in terrain modelling. **ITC Journal**, 1985-2:101-106.

Frederiksen, P., Jacobi, O. & Kubik, K., 1986. Optimum sampling spacing in digital elevation models. **International Archives of Photogrammetry and Remote Sensing,** 26(3/1):252-259.

Fritsch, D., 1984. Proposal for the determination of the least sampling interval for DEM data acquisition. **Presented Paper, Digital Elevation Models Workshop,** Edmonton, Canada.

Fritsch, D., 1988. Some experience with the determination of optimum sampling density. **International Archives of Photogrammetry and Remote Sensing,** 27(B11):III-493-504.

Frodesen, A., Skjeggestad, O. & Tofte, H., 1979. **Probability and Statistics in Particle Physics.** Universitetsforlaget, Oslo, 501pp.

Gannapathy, S. & Dennehy, T., 1982. A new general triangulation method for planar contours. **Computer Graphics,** 16(3): 69-72.

Gar-Al-Nabi, I., 1988. **An Examination of the Comparitive Accuracy of Digital Terrain Models Derived from Data Captured by Cartographic Digitising of Contours at Various Scales.** M. App. Sci. Thesis. University of Glasgow, 176pp.

Green, R. & Sibson, R., 1978. Computing Dirichlet tessellations in the plane. **The Computer Journal,** 22(2):168-173.

Gut, D. & Hohle, J., 1977. High altitude photography: Aspects and results. **Photogrammetric Engineering and Remote Sensing,** 43(10):1245-1255.

Hallert, B., et al., 1968. Quality problems in photogrammetry. **International Archives of Photogrammetry,** 17(3).

Hannah, M., 1981. Error detection and correction in digital terrain models. **Photogrammetric Engineering and Remote Sensing,** 47(1):63-69.

Harley, J., 1975. **Ordnance Survey Maps: A Descriptive Mamual.** Ordnance Survey, Southampton. 200pp.

Haywood, P., 1981., **An Estimation of the Size of a Small Topographic Data Base of Great Britain and Some Considerations Arising.** M.App.Sci. Thesis. University of Glasgow.

Hobbie, D., 1977. C-100 Planicomp, the analytical stereo-plotting system from Carl Zeiss. **Photogrammetric Engineering and Remote Sensing,** 43(11):1377-1390.

Hogg. R. & Tanis, E., 1977. **Probability and Statistical Inference.** Macmillan Publishing Co., Inc. N.Y. 450pp.

Imhof, E., 1965. **Kartographische Gelandedarstellung.** Walter de Gruyter & Co., Berlin. 425pp.

Inaba, K., Aumann, G. & Ebner, H., 1988. DTM generation from digital contour data using aspect information. **International Archives of Photogrammetry and Remote Sensing,** 27(B9):III-101-110.

Kelk, B., 1973. **Hardware for Automatic Cartography.** in "Automation in Cartography". British Cartographic Society. 25-50.

Konecny, G., Bahr, H., Reil, W. & Schreiber, H., 1979. **Use of Spaceborne Metric Camera for Cartographic Applications.** Report to the Ministry of Research and Technology of FRG.

Koppe, C., 1902. Ueber die Zweckentsprechende Genauigkeit der Hoehendarstellung in topographischen plaenen und karten fuer allgemeine technische Vorarbeiten. **ZFV,** 34(2):33-38.

Koppe, C., 1905. Die Neue topographischen landeskarte des herzogtums braunschweig in massstable 1:10,000. **ZFV,** 36(14):397-424.

Kubik, K. & Botman, A., 1976. Interpolation accuracy for topographic and geological surfaces. **ITC Journal,** 1976(2):236-274.

Kubik, K. & Roy, B., 1986. **Digital Terrain Model Workshop Proceedings,** Columbus, Ohio. 1986. 150pp.

Laiho, A. & Kilpela, E., 1988. Stability testing of analytical plotter. **International Archives of Photogrammetry and Remote Sensing**, 27(B10):II-280-289.

Leberl, F. & Olson, D., 1982. Raster scanning for operational digitizing of graphical data. **Photogrammetric Engineering and Remote Sensing**, 48(4):615-627.

Ley, R., 1986. Accuracy assessment of digital terrain models. **Auto Carto London**, 1:455-464.

Ley, R., 1990. A DTM of Great Britain based on 1:50,000 scale Ordnance Survey series - A production view point. **IBG90 (Annual Conference)**, Terrain modelling session. University of Glasgow. Jan., 1990.

Li, Q., 1982. **Engineering Surveying** (text book). Publishing House for Surveying and Mapping. Beijing. (in Chinese).

Li, Z., 1988a. On the measure of digital terrain model accuracy. **Photogrammetric Record**, 12(72):873-877.

Li, Z., 1988b. An algorithm for compressing digital contour data. **The Cartographic Journal**, 25(2):143-146.

Linderberg, J., 1986. ARIMA processes for modelling digital terrain profiles. **International Archives of Photogrammetry and Remote Sensing**, 26(3/2):427-441.

Lindgren, R., 1983. The establishment of ground truth for the ISPRS Working Group III:3. **Proceedings of International Colloquium on Mathematical Aspects of DEM.** Stockholm.

Lindig, G., 1956. Neue Methoden der Schichtlinienprufung. **ZFV**, 1956-7:244-296.

Lowthian, B., 1986. **The Accuracy of a DTM Derived from O.S. 1:50,000 Scale Mapping.** Unpublished diploma project. University of Glasgow.

Makarovic, B., 1972. Information transfer in construction of data from sampled points. **Photogrammetria**, 28(4):111-130.

Makarovic, B., 1973. Progressive sampling for DTMs. **ITC Journal**, 1973-4:397-416.

Makarovic, B., 1974. Conversion of fidelity into accuracy. **ITC Journal**, 1974-4:506-517.

Makarovic, B., 1975. Amended strategy for progressive sampling. **ITC Journal**, 1975-1:117-128.

Makarovic, B., 1977. Composite sampling for DTMs. **ITC Journal**, 1977-3:406-433.

Makarovic, B., 1979. From progressive sampling to composite sampling. **Geo-processing**, 1:145-166.

Makarovic, B., 1984. Selective sampling for digital terrain modelling. **International Archives of Photogrammetry and Remote Sensing**, 24(4a):264-277.

Mandelbrot, B., 1976. How long is the coast of Britain ? Statistical self-similarity and fractional dimension. **Science**, 155:636-38.

Mandelbrot, B., 1982. **The Fractal Geometry of Nature.** W. H. Freeman and Company. San Francisco. 460pp.

Mark, D., 1975. Geomorphometric parameters: a review and evaluation. **Geografiska Annaler**, 57A:165-177.

McCullagh, M. & Ross, R., 1980. Delaunay triangulation of a random data set for isarithmic mapping. **The Cartographic Journal**, 17(2):178-181.

McCullagh, M., 1983. "If you're sitting comfortably we'll begin". **Workshop Notes on Terrain Modelling**, Australian Computing Society, Siren Systems, 158pp.

McLain, D., 1976. Two dimensional interpolation from random data. **The Computer Journal**, 19(2):178-181.

McMaster, P., Haywood, P. & Sowton, M., 1986. Digital mapping at Ordnance Survey. **Auto Carto London**, 1:13-23.

Meyer, W., 1985. **Concepts of Mathematical Modelling.** McGraw-Hill Book Company. 440pp.

Mikhail, E., 1976. **Observations and Least Squares.** IEP, New York. 497pp.

Miller, C. & Laflamme, R., 1958. The digital terrain model - theory and applications. **Photogrammetric Engineering,** 24:433-442.

Mirante, A. & Weingarten, N., 1982. The radial sweep algorithm for constructing triangulated irregular networks. **IEEE Computer Graphics and Application,** 2:11-21.

Mitchell, C., 1973. **Terrain Evaluation.** Longman, London. 221pp.

Ostman, A., 1986a. A graphic editor for digital elevation models. **Geo-processing,** 3:143-154.

Ostman, A., 1986b. A PC-based editor for digital terrain models. **Auto Carto London,** 1:465-474.

Ostman, A., 1987. Accuracy estimation of digital elevation data banks. **Photogrammetric Engineering and Remote Sensing,** 53(4):425-430.

Petrie, G., 1987a. Photogrammetric methods of data acqusition for terrain modelling. **Terrain Modelling in Surveying and Civil Engineering:** A Short Course. Universities of Surrey and Glasgow.

Petrie, G., 1987b. Terrain data acquisition and modelling from existing maps. **Terrain Modelling in Surveying and Civil Engineering:** A Short Course. Universities of Surrey and Glasgow.

Petrie, G., 1990. **Analogue, analytical and digital photogrammetric systems applied to aerial mapping.** in "Engineering Surveying Technology", edited by T. Kennie & G. Petrie. 238-288.

Petrie, G. & Kennie, T., 1986. Terrain modelling in surveying and civil engineering. **Conf. on 'State of The Art' in Stereo and Terrain Modelling.** British Computer Society, Display Group. Also **CAD Journal,** 19(4):171-187.

Petrie, G. & Kennie,T., 1987. An introduction to terrain modelling: applications and terminology. **Terrain Modelling in Surveying and Civil Engineering:** A Short Course. Universities of Surrey and Glasgow.

Peucker, T., 1972. **Computer Cartography.** Assoc. of Amer. Geog., Comm. on College Geog. Washington, D.C. 75pp.

Reinhardt, W., 1986. Optical superimposition of stereo model and graphical information as a tool for DTM quality control. **International Archives of Photogrammetry and Remote Sensing**, 26(4):207-215.

Reinhardt, W., 1988. On line generation and verification of digital terrain models. **International Archives of Photogrammetry and Remote Sensing**, 27(B11):III-546-555.

Richardus, P., 1973. The precision of contour intervals of large and medium scale maps. **Photogrammetria**, 29:81-107.

Rinner, K. & Burkhardt, R., 1972. **Photogrammetrie.** Jordan/Eggert/ Kneissl Handbuch der Vermessungskunde, Band IIIa. J.B.Metzlersche Verlagsbuchhandlung. Stuttgart.

Rivett, P., 1972. **Principles of Model Building:** The construction of Models for Decision Analysis. John Wiley & Sons. 141pp.

Roberts, R. 1957. Using new methods in highway location. **Photogrammetric Engineering**, 23:563-569.

Rollin, J., 1986. A method of assessing the accuracy of cartographic digitising tables. **The Cartographic Journal**, 23(2):144-146.

Saaty, T. & Alexander, J., 1981. **Thinking with Models.** Pergamon Press. 181pp.

Savolainen, A. & Ruotsalainen, R., 1976. Experiences of the ISP standard tests. **The Photogrammetric Journal of Finland**, 7(1):17-32.

Schult, R., 1974. A system for digital terrain models. **Allgemeine Vermessungs-Nachrichten**, Nr.8.

Schut, G., 1976. Review of interpolation methods for digital terrain models. **International Archives of Photogrammetry**, 21(3).

Schwarz, P., 1982. A test for personal stereoscopic measuring precision. **Photogrammetric Engineering and Remote Sensing**, 48(3):375-381.

Shearer, J., 1987. The accuracy of digital terrain models. **Terrain Modelling in Surveying and Civil Engineering**: A Short Course. Universities of Surrey and Glasgow.

Shearer, J. & Li, Z., 1990. The accuracy of digital terrain models. **IBG90 (Annual Conference)**, Terrain modelling session. University of Glasgow. January 1990.

Sigle, M., 1984. A digital terrain model for the state of Baden-Wurttemberg. **International Archives of Photogrammetry and Remote Sensing**, 25(A3/B):1016-1026.

Stamp, L. & Beaver, S., 1971. **The British Isles.** 6th edtion. Longman. 881pp.

Stark, E., 1976. The effect of angular field on horizontal and vertical accuracy in photogrammetric plotting. **Proceedings of 35th Photogrammetric Week.** University of Stuttgart: 129-144.

Strahler, A., 1956. Quantitative slope analysis. **Bulletin of the Geological Society of America**, 67:571-596.

Tempfli, K., 1980. Spectra analysis of terrain relief for the accuracy estimation of digital terrain models. **ITC Journal**, 1980-3:478-510.

Tempfli, K. & Makarovic, B., 1979. Transfer functions of interpolation methods. **Geo-processing**, 1:1-26.

Tobler, W., 1969. Geographical filters and their inverses. **Geographical Analysis**, 1:234-253.

Togliatti, G. & Moriondo, A., 1985. Evaluation of the accuracy of the the Metric Camera images for the production of line maps and orthophotos. **Metric Camera Workshop**, 29-34.

Toomey, M., 1984. **Digital Elevation Model Workshop Proceedings.** Edmonton, Alberta. 231pp.

Toomey, M., 1988. The Alberta digital elevation model. **International Archives of Photogrammetry and Remote Sensing,** 27(B3):775-783.

Torlegard, K., Ostman, A. & Lindgren, R., 1986. A comparative test of photogrammetrically sampled digital elevation models. **Photogrammetria,** 41(1):1-16.

Tuladhar, A. & Makarovic, B., 1988. Digital terrain models (DTM) from contour lines upgraded by photogrammetric selective sampling. **International Archives of Photogrammetry and Remote Sensing,** 27(B8):IV-94-103.

Turner, H., 1977. A comparison of some methods of slope measurement from large-scale air photos. **Photogrammetria,** 32:209-237.

Uffenkamp, D., 1986. Improvement of digital mapping with graphic image superimposition. **International Archives of Photogrammetry and Remote Sensing,** 27(3/2):665-671.

Warwick, G., 1964. **Relief and Structure.** in "The British Isles: A Systematic Geography",edited by J. Watson. Nelson, London and Edinburgh. 91-109.

Wentworth, C., 1930. A simplified method of determining the average slope of land surface. **American Journal of Science,** 20:184-194.

Wild, E., 1980. Interpolation with weight-functions -- A general interpolation method. **International Archives of Photogrammetry and Remote Sensing,** 23(B3):780-793.

Yoeli, P., 1975. **Compilation of Data for Computer-assisted Relief Cartography,** in "Display and Analysis of Spatial Data". edited by J. Davis and M. McCullagh. John Wiley and Sons. 352-367.

Yoeli, P., 1977. Computer executed interpolation of contours into arrays of randomly distributed height points. **The Cartographic Journal**, 14:103-108.

Yoeli, P., 1983. About cartographic contouring with computers. **Auto Carto VI**, 262-266.

Ziegenfus, R., 1981. An application of computer mapping in depicting spatial and temporal cpmpoments of cancer mortality. **Computer Mapping Applications in Urban, State and Federal Government.** Harvard Library of Computer Graphics, 16:133-140.