Modelling, Interpretation and Inversion of Multielectrode Resistivity Survey Data

By

Panagiotis Tsourlos

Thesis submitted to the

Department of Electronics

and the committee on graduate studies of the

University of York

in partial fulfilment of the requirements

for the Degree of

Doctor of Philosophy

November 1995

To my parents

Στους γονείς μου

Ιωάννη και Θεοδότα

Abstract

This thesis studies modelling and 2-D inversion schemes for earth resistivity data. Initially a detailed study of the basic theory, mechanisms and instrumentation of the resistivity surveying technique is presented.

The Finite Element technique was chosen to perform the forward modelling. The features and accuracy of the 2.5-D FEM modelling scheme which was developed for that work are discussed. The FEM technique was modified to accommodate the calculation of arrays parallel to the strike direction. Further, the FEM is used to study the effect of the terrain topography on commonly used resistivity arrays. The results indicate that terrain can be a significant source of noise for all of the arrays.

The general principles, of the resistivity inverse problem are presented. Ways for calculating the Jacobian matrix (within the FEM context) are presented. The adjoint equation technique was found to be both accurate and computationally economical. The use of the Jacobian matrix to explain unusual apparent resistivity responses and to design optimum surveys is demonstrated.

The basic principles merits and demerits of commonly used "approximate" techniques for reconstructing earth resistivity data are discussed. Further, a generalized (iterative) back-projection algorithm is proposed. The merits and limitations of the algorithm are discussed and examples of its performance with synthetic and real data are given.

Further, widely used "accurate" inversion techniques are studied. Some of the advantages and limitations of the techniques are demonstrated by means of synthetic examples. Finally, a fast smoothness constrained inversion algorithm which uses a quasi-Newton technique for updating the Jacobian matrix is proposed and its features are fully explained. Extensive tests of the algorithm with synthetic and real data indicate that it is a reliable and robust tool for data interpretation.

Acknowledgments

First of all I would like to express my gratitude to my supervisor Dr. John Szymanski for his constant guidance, support and encouragement throughout my time at York and also for helping me obtain additional funding during the last two years of this project. I would like to thank Dr. Julian Tealby for his support and advice during the past years. I acknowledge the members of the Remote Sensing laboratory: my good friend Dr. Jonathan Dittmer in particular for introducing me into the Finite Element method and making his code available to me. Jon's advice and help contributed most in this project.

I thank the members and students of the Geophysical laboratory in the University of Thessaloniki for helping me collect some of the data presented in this work. I am particularly grateful to Dr. Gregory Tsokas for his support and advice and for helping me obtain some extra funding. Thanks are due to the British School at Athens and the British Geological Survey (Dr. R. Ogilvy in particular) for partially funding this project. I also thank the Advanced Geoscience Inc. for providing me with the Sting Cave data.

I would like to thank the many good friends I have made during my staying in York as well as my friends in Greece. I thank them all for their friendship, support and solidarity. I am particularly grateful to Faye Karavia: her moral support during the last two years was invaluable. I would also like to thank my best friend and colleague Antonis Giannopoulos for sharing with me all the joys and frustrations involved in the doctorate work. During these difficult years Antonis was always there for me and I am grateful for that.

Most of all I am deeply grateful to my family and particularly my parents. Without their love and unconditional support this project would not be possible. The least I can do is to dedicate this work to them.

Declaration

This thesis is the work of the author except where indicated. Work contained in this thesis has previously been presented as follows:

Tsourlos, P., Dittmer, J. and Szymanski, J. (1995). A study of non-linear techniques for the 2-D inversion of earth resistivity data. Expanded abstracts of the 57th meeting and technical exhibition of the EAEG: Glasgow, Scotland, 29 May- 2 June, 1995, P031.

Tsourlos P., Szymanski J., Ogilvy R. and Jackson P. (1995). Inversion of earth resistivity data from the multi-electrode RESCAN system. Expanded abstracts of the 57th meeting and technical exhibition of the EAEG: Glasgow, Scotland, 29 May- 2 June, 1995, P027.

Tsourlos P., Szymanski J., and Dittmer J. (1995). The topographical effect in earth resistivity arrays: a comparative study. Proceedings of the IEEE International Geoscience and Remote Sensing Symposium: Florence, Italy, 10-14 July, 1995, Vol 1, 30-34.

Tsourlos P., Dittmer J. and Szymanski J. (1995). Non-linear techniques for the inversion of earth resistivity data Paper presented at the IEEE International Geoscience and Remote Sensing Symposium: Florence, Italy, 10-14 July, 1995.

Tsourlos P., and Szymanski J. (1995). Self-correcting approaches to the inversion of resistive tomography field survey data. Poster presented at the Archaeological Prospection Conference, University of Bradford, U.K., 12-13 September, 1995.

Tsourlos P. and Szymanski J. (1995). The impact of topographical variations on resistivity surveys: A quantitative study. Poster presented at the Archaeological Prospection Conference, University of Bradford, U.K., 12-13 September, 1995.

Tsokas G., Giannopoulos A., Tsourlos P., Vargemezis J., Tealby J., Sarris A., Papazachos C. and Savopoulou T. (1994). A large scale geophysical survey

in the archaeological site of Europos (N.Greece). Journal of Applied Geophysics, 32, 85-98.

Szymanski J., Tsourlos P. and Dittmer J. (1994). Maximum entropy reconstruction of resistivity data sets using a finite element forward model. Expanded abstracts of the 56th meeting and technical exhibition of the EAEG: Vienna, Austria, 6-10 June, 1994, P128.

Tsokas G., Tsourlos P. and Szymanski J. (1994). Square array resistivity anomalies and inhomogeneity ratio calculated by the finite element method. (submitted to Geophysics).

Dittmer J., Tsourlos P. and Szymanski J. (1994). A discussion of several aspects of the Finite Element method and its use in the resistivity modelling of the subsurface. Expanded abstracts of the 56th meeting and technical exhibition of the EAEG: Vienna, Austria, 6-10 June, 1994, P127.

Szymanski J. and Tsourlos P. (1993) The resistive tomography technique for archaeology: an introduction and review. Archaeologia Polona 31, 5-31.

Szymanski J. and Tsourlos P. (1993). Inversion of resistivity data sets arising from electrical tomography. Expanded abstracts of the 55th meeting and technical exhibition of the EAEG: Stavanger, Norway, 7-11 June, 1994, P127.

Tsourlos P., Szymanski J., Dittmer J. and Tsokas G. (1993). The use of back-projection for fast inversion of 2-D resistivity data. Proceedings of the 2nd congress of the Greek Geophysical Union, Florina, Greece, 5-7 May, 1993, 1, 71-81.

Tsourlos P. (1993). Modelling and inversion schemes for vertical profiling resistivity data. Internal report, Department of Electronics, University of York.

Coppack P., Emerick K., Wilson K., Dittmer J, Szymanski J., Tsourlos P. and Giannopoulos A. (1992). Recent archaeological discoveries at the medieval site of Fountains abbey. Proceedings of MEDIEVAL EUROPE 1992, York, U.K. 21-24, September 1992, York, UK, 201-206.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Thesis subject and aims

Geophysical techniques are well-established and widely used to solve a variety of subsurface detection problems. This study will deal with one particular technique which involves the measurement of the variations of the earth's resistivity. As in any geophysical technique, these measurements do not provide a direct "image" of the subsurface but simply the integrated effect of the subsurface property which could be (in cases of complex subsurface property distribution) far removed from reality.

This work deals with inversion schemes concerning earth-resistivity data sets. The term "inversion" in the resistivity method describes the (usually fully computerized) procedure of constructing an image of the subsurface's resistivity distribution given the respective observed data sets. There are two general approaches to the resistivity inverse problem: "approximate" schemes which, despite their theoretical weaknesses, can produce useful images of the subsurface in a limited time, and "accurate" schemes which are more correct as far as their theoretical assumptions are concerned but commonly involve heavy computational loads. This study will deal with both approaches. For computational reasons the inversion will be restricted to

2-D schemes.

From the beginning of the project it became clear that a full study of inversion algorithms cannot be achieved without addressing the resistivity forward modelling procedure first (forward modelling is the procedure of calculating the observed data set given a model property distribution): the modelling procedure is vital for understanding the nature of the resistivity problem and, moreover, can be used as a platform for developing efficient interpretation algorithms.

Both resistivity modelling and 2-D inversion procedures have been widely studied over the last decades, so the following question has to be answered as far as this project is concerned: is there a real need for further studies in this area? This question is answered below.

The need for this project

Resistivity offers several relative advantages when compared to other geophysical techniques (low survey cost, simple to implement etc.). On the other hand, its main disadvantage is that it is traditionally considered to be a low resolution technique, mainly due to hardware limitations.

In the last few years, the advent of resistivity measuring instruments which allowed the full automation of the measuring procedure has a significant impact on the way resistivity is being viewed: the new instruments allow the collection of a large number of measurements in a relatively short time and, hence, resistivity cannot be considered to be a low resolution technique any more. But most of the existing interpretation techniques cannot cope with this increased amount and type of measurements.

Therefore it is clear that there is scope for reviewing the existing resistivity inversion techniques, and suggesting possible ways for improving them in view of the new automated measuring procedure. Although the algorithms presented should be able

to cope with any problem, regardless of its size, the examples and cases examined by this work will be from shallow depth targets. [1]

General guidelines and approach

The general guidelines and approach that were followed throughout this project are as follows:

- A detailed study of the basic theory, mechanisms and instrumentation of the resistivity surveying technique is necessary before proceeding into examining modelling and inversion schemes.

- The existing forward modelling procedures have to be examined. The forward modelling scheme which will be chosen should be:
  a) able to model any resistivity structure (no matters how complicated).
  b) able to cope with circumstances encountered in surveys (i.e. topography).
  c) flexible enough to cope with any measuring scheme.
  d) able to calculate quantities that are needed by the interpretation/inversion schemes. (i.e. the Jacobian matrix).

- The existing "approximate" and "accurate" inversion algorithms should be critically studied before any new algorithms are proposed. Bearing in mind the automated measuring procedure, the proposed inversion schemes should

[1] At this point it is necessary to discuss what is meant by the term "shallow" depth targets. There are no general limits for what is shallow or deep in geophysics. It is always dependent on the kind of features which are sought. For instance, 5 meters most of the time, is considered to be very deep for archaeological features, while it is very shallow for water surveys. At the same time 300 meters is deep for water exploration, while it is quite shallow for oil prospection. On the other hand, every geophysical method has, in respect of the penetration depth, physical and instrumentation limitations that must be considered. Taking all of this into account, it is believed that for the type of measurements taken by an automated measuring scheme the term "shallow" corresponds to an approximate depth between 0-150 meters. The kind of targets appropriate to these depths arise in : archaeological, civil engineering, environmental and hydrogeological applications.

have the following characteristics:

a) they should be flexible enough to cope with any measuring scheme.

b) They should be relatively fast, since they have to cope with an increased number of measurements.

Other general constraints that were set are:

- the thesis should be in a form that is accessible by a broad audience since geophysical techniques are being applied in a variety of disciplines (geophysicists, geologists, engineers, archaeometrists etc.).

- The proposed schemes should be fully automated and thus employable on a computer system, preferably a desktop IBM PC compatible. But, to allow further flexibility (and since the computational needs may vary depending on the size of the problem) the programs should be readily transferable to any type of computer.

- The algorithms should be reliable: extended testing has to be performed under a controlled environment. This can be achieved by the use of synthetic data.

- Most importantly, since the entire project is trying to address a "real" need the algorithms should be able to cope well with real data. Thus, extensive tests with real data from areas of known conditions have to be performed in order to really evaluate the algorithms.

1.2 The Structure of the Thesis

The structure of the thesis reflects the structured approached mentioned above:

Chapter 2 gives the basic background regarding the earth-resistivity method. This chapter provides the introductory material useful for the justification and understanding of the work presented in the following chapters and focuses on

some aspects of the resistivity method which are not very clear in the current literature. The basic background regarding the conduction of the current flow into the earth is presented and basic concepts in resistivity prospecting are explained. The resistivity arrays and the resistivity measuring modes are described and a comparison of the performance of widely used arrays is discussed. The instrumentation used for earth resistivity measurements is described. Finally, the field application procedures for a resistivity survey are presented in detail.

Chapter 3 deals with the forward resistivity modelling. The existing approaches for the solution of the forward resistivity problem are discussed and among them the finite element method (FEM) is chosen for this work. Although a detailed analysis of the application of the 2.5-D FEM into the resistivity modelling is given in many works, the scheme developed for this work is described in full. This is necessary because full justification of techniques that are presented in this work is possible only if the core of the 2.5-D FEM modelling is fully explained. The computational aspects and accuracy of the proposed scheme are addressed. Further, a strategy for modelling arrays parallel to the strike direction (i.e. square array) is proposed. Finally the scheme is used to study the effects of terrain topography on commonly used resistivity arrays.

Chapter 4 In the first part of this chapter the general concept of the inversion procedure is discussed, with a view to the application of the inversion schemes to the resistivity data. The theoretical and practical limitations of the inversion procedure are discussed and the general usefulness of the inversion is addressed. Further, general approximate and accurate ways for treating the resistivity inverse problem are presented. It will be shown that the Jacobian matrix is crucial in both approximate and accurate schemes. Therefore, before attempting to describe the inversion schemes analytically ways for calculating the Jacobian matrix are presented in detail. Particular emphasis is given to the application of those techniques within the finite element framework. The

accuracy and overall performance of the tested techniques is checked. One of those techniques is used to calculate the sensitivity of commonly used arrays. It will be shown that the sensitivity matrix can be used to justify the response of the arrays and to design optimum resistivity surveys.

Chapter 5 presents approximate inversion schemes, which, despite their intrinsic theoretical weaknesses, can produce reasonably valid sectional images of the subsurface resistivity pattern in a limited time. The need for fast algorithms is increased by the development of the automatic measuring systems. The most commonly used approximate inversion algorithms will be reviewed (pseudo-section, Barker's method, back-projection etc.). Medical imaging algorithms will be presented as well. An explanation of why these techniques work is given by means of the Jacobian matrix. Finally, a generalized (iterative) back-projection algorithm is proposed. The need for such an algorithm derives from the analysis of the limitations of the existing approximate algorithms. The algorithm can include all of the presented BP algorithms as well as techniques such as the pseudosection and Barker's method. Secondly it can reconstruct data from any array as well as from unconventional probe arrangements. This is achieved by recognizing the major importance of the Jacobian matrix within the approximate reconstruction procedure. The merits and limitations of the algorithm are discussed and examples of its performance with synthetic and real data are given.

Chapter 6 presents non-linear inversion techniques applied to the 2-D reconstruction of earth resistivity data. In the first part of this chapter widely used techniques such as the non-linear least-squares method, Marquadt's method, and smoothness constrained (Occam's) inversion will be reviewed. Further, approaches such as inversion with prior information, simulated annealing and maximum entropy will be presented as well. The advantages and limitations of the techniques will be demonstrated by means of synthetic examples. In the second part of this chapter a fast smoothness constrained inversion algo-

rithm which uses a quasi-Newton technique for updating the Jacobian matrix is proposed. The need for such an algorithm is discussed. The features of the algorithm are presented in detail and comparisons to other techniques will be shown. Finally, extensive tests of the algorithm with real data will be presented.

Chapter 7 presents a summary of the main points being made in this thesis. Further possible improvements and lines of research are discussed.

Chapter 2

Resistivity Background

In this chapter the basic background regarding the earth-resistivity method is presented. This chapter is not meant to be a detailed review of the basic theory (comprehensive reviews can be found in any exploration geophysics text-book) but is written with the intention:

a) to provide the introductory material useful for the justification and understanding of the work presented in the following chapters.

b) to focus on some aspects of the resistivity method which are not very clear in the current text book literature.

In this chapter the basic background regarding the conduction of the current flow into the earth is presented and basic concepts in resistivity prospecting are explained. The resistivity arrays and the resistivity measuring modes are described and a comparison of the performance of widely used arrays is discussed. The field application procedures for a resistivity survey are presented as well. Finally, the instrumentation used for earth-resistivity measurements is described and some problems commonly encountered in resistivity prospecting are discussed.

2.1 Basic Theory

2.1.1 Resistivity and Geophysics

Several geophysical techniques exist for the detection of "shallow" subsurface features. Among them the most frequently used are: resistivity, induced polarization, electromagnetic methods, ground probing radar and magnetometery. The first four methods are described as "active" since the effect of an artificially created field is being measured, while magnetometry is considered to be a "passive" method because the effect of an existing field is being measured. A short description of the techniques and their most common applications follows :

Resistivity method With this technique the variations of the earth's resistivity are measured. An electrical current is introduced into the ground and the resultant measured potential differences at the surface give an indirect indication of the subsurface resistivity distribution.

Induced polarization method this is very similar to the resistivity method but it makes use of the capacitance actions which are caused by conductive bodies. It is used mainly for mineral exploration (Sumner, 1976; Pelton et al., 1978) and for archaeological exploration (Aspinall and Lynam, 1970) as well.

Electromagnetic methods An alternating EM field (primary) is introduced into the ground. This field induces eddy currents in an existing conductive body which then becomes a source of a secondary EM field. The resulting effect of these two fields is measured and in this way the electric and magnetic properties of the body can be obtained. Electromagnetic methods are the basic techniques used for the detailed mapping of mineral resources (Ward and Hohmann, 1988) but are also being used in archaeological prospection (Clark, 1990).

Ground Penetrating Radar method Electromagnetic waves are coupled into the ground through a transmitting antenna. These waves are partly reflected at any dielectric discontinuity. The reflected waves are received back and provide information about the discontinuity. Radar is being used extensively in civil engineering surveys (Daniels et al., 1988) and is also being used in archaeological surveys (Vaughan, 1986)

Magnetometery The variations of the earth's magnetic field are measured and in this way targets of high (or low) magnetic susceptibility can be located. Magnetometery is used for initial mineral exploration and is also used for geophysical surveys in archaeology (Weymouth, 1986; Tsokas and Papazachos, 1992).

In general, there is no single optimum method which can always give valid and useful results independent of the target characteristics, since every method tends to pick up some types of targets better than others. Furthermore, it is not only the target characteristics that are crucial, but also the local environment. This is because in geophysical surveys it is not necessarily the actual value of a property that is of primary interest, but rather the difference between the property value of the feature compared to the surrounding environment. Therefore, it is the responsibility of the person who carries out the survey to choose the method which could give the best results, judging by the local field conditions. The decision is made considering several other, mainly financial, factors such as the instrumentation cost and the limitations of time, money and effort for the survey to be conducted, and finally the available processing facilities.

Theoretically, the best geophysical survey would have been the one for which as many methods as possible were used, since the maximum amount of information could be collected. In practice, this is not possible, so usually just two (or sometimes even one) methods are felt to be enough to provide an adequate image of the subsurface.

This work will be limited to the resistivity method and, in particular, to the vertical profiling resistivity techniques. Compared to the other methods resistivity has some general advantages:

- The instruments for resistivity surveys (Resistivity meters) cost 2 to 20 times less than these used in other techniques, making the implementation of resistivity methods financially attractive.
- Resistivity results are relatively easy to interpret (especially resistivity profiling) and even without processing can give reasonably valid subsurface insights; but this happens only when the features have simple structures.
- Resistivity can provide information in both the lateral and vertical directions.
- Resistivity is not particularly sensitive to external factors such as power cables or metallic litter.

On the other hand the most significant limitations of resistivity are the following:

- Complex structures render the resistivity interpretation rather complicated.
- It is not entirely nondestructive, since probes have to be stuck into the area under prospection. Therefore, it is not, readily applicable to surveys in built constructions.
- Data acquisition can become quite complicated and laborious when large number of measurements are to be acquired. Furthermore, the method's penetration depth is limited by hardware factors (cable length, electrical power).
- It is sensitive to ground moisture, topography changes and surface resistivity variations

As is obvious, resistivity is not the "perfect" method. Despite its disadvantages, however, it can provide high quality results and in several cases could give adequate information about the subsurface without the need to apply other techniques.

Applications

The literature contains numerous examples of successful application of the resistivity technique to a variety of subsurface problems. The most common applications of resistivity are as follows:

Geological Resistivity has been extensively used for geological mapping (Vandenberghe, 1982; Olesen et al., 1992; Griffiths and Barker, 1993)

Hydrogeological This is probably the most popular application of the resistivity technique. Resistivity has been used for hydrogeological mapping and groundwater exploration (Van dam, 1976; Rijo et al., 1977; Aubert et al., 1984; Olayinka and Barker, 1990).

Geothermal Resistivity has been successfully used in geothermal field exploration (Tripp et al., 1977; Wright et al. 1985; Thanassoulas and Tsokas, 1987).

Engineering Resistivity is being used to a variety of engineering problems such as the foundations and integrity of dams, cavity detection, planning of infrastructure, assessing the hydraulic and anisotropical properties of the subsurface, etc. (Habberjam, 1975; Smith, 1986; Butler and Llopis, 1990; Dahlin et al. 1994).

Environmental Several applications of resistivity have been reported in environmental problems such as detecting ground-water pollution (Rogers and Kean, 1981; Urish, 1983), landfill management (Barker, 1992) monitoring contaminant leakage (Van et al., 1992), etc.

Archaeological Resistivity is the most popular geophysical technique used in archaeometry and has been used for decades to locate a variety of targets of archaeological interest (Aitken, 1974; Hesse et al., 1986; Roka and Tsokas, 1987; Orlando et al. 1987; Szymanski et al., 1992).

2.1.2 Conduction of electricity in soil

The resistivity method involves the introduction of direct electrical current into the ground. There are three ways via which the electrical current can be conducted within the earth:

Electrolytic conduction: The electrical current propagates through the pores of the rocks or the soil which are filled with water, containing ions of dissolved salts and minerals. This is actually an ionic conduction and it is extremely important for the resistivity method because most rocks conduct electrical current with this process.

Ohmic conduction: The electrical current propagates via the crystalline structure of some materials, mainly metals. This kind of conduction, also known as electronic, is quite important for mineral exploration.

Dielectric conduction: The existence of an alternating electrical field can cause ions in the structure of insulating materials to have a cyclic change in their positions. This movement can be detected as an alternating current. Despite the fact that AC current is used in resistivity prospection, it is of such a low frequency that dielectric conduction is usually considered to be negligible.

The goal of the resistivity method is to measure the potential differences on the surface due to the current flow within the ground. The measured drop of potential reflects the difficulty with which the electrical current can be made to flow through the earth, giving an indication of the earth's electrical resistivity ρ, which is directly dependent on the way the current is being conducted within the earth. Since current conduction is related to the lithology (ohmic conduction) and ground water (electrolytic conduction) of the subsurface, a knowledge of resistivity can be the basis for distinguishing existing earth features (layering, structures). The electrical resistivity ρ of a cylindrical solid of length L and cross section A having resistance R between the end faces is given by:

ROCK TYPE	RESISTIVITY RANGE (Ωm)
IGNEOUS	$10^2 - 10^6$
LIMESTONE	$10^2 - 10^4$
SANDSTONE	$10^2 - 10^3$
SOIL	$10^0 - 10^1$
MINERALS	$10^{-8} - 10^0$

Table 2.1: Typical ranges of resistivity for some characteristic types of rocks (after Aitken, 1974)

$$\rho = \frac{R.A}{L} \tag{2.1}$$

Where R is in ohms, L is in metres and A is in square metres. The unit of resistivity, ρ, is the ohm-meter (Ωm). Another term for describing the earth's behaviour towards current flow is conductivity, σ, which is the reciprocal of resistivity $\sigma = 1/\rho$, and thus physically represents the ease with which the electric current can be made to flow through the earth. The unit of conductivity is siemens per meter (S/m). Typical ranges of resistivity for some characteristic types of rocks are shown in Table 2.1.

It should be stressed at this point that the effect of electrolytic conduction is the most important in governing the earth's resistivity. The factors that determine electrolytic conduction in the ground (and therefore the resistivity) are a combination of considerations concerning water content and water composition of the subsurface. According to (McNeill, 1980) and (Tagg, 1964) the most important factors are :

a) The water distribution, which depends on weather, time of the year, depth of water table, type of soil.

b) Chemical composition and concentration of the salts contained within the water.

c) The grain size of the soil and the pore size of the rocks, combined with any possible fractures.

ROCK TYPE	% WATER	RESISTIVITY (Ωm)
SANDSTONE	0.10	1.4 10^8
SANDSTONE	1.00	4.2 10^3
GRANITE	0.00	1.0 10^{10}
GRANITE	0.19	1.8 10^6
GRANITE	0.31	4.4 10^3

Table 2.2: The variation of resistivity of some rocks with respect to the water content (after Telford et al., 1990).

d) The temperature: earth resistivity highly depends on the water present and water's resistivity is known to have a large temperature dependence (Tagg, 1964). The variation of resistivity of some rocks with respect to the water content is shown in Table 2.2. Since the factors deciding electrolytic conduction are very variable, it is often observed that similar physical formations can appear as entirely different variations in resistivity. This fact renders resistivity as a property quite unstable, and hence sometimes inadequate, as a tool for using it to extract exact lithological conclusions for the subsurface. Therefore it should be born in mind while interpreting resistivity data that the measured resistivity values are not absolute but relative, and therefore only relative conclusions about the area's lithology can be made.

For instance, by observing some data it may be said that there is a formation that is less resistive than the surrounding formations, but it will not be safe to decide its lithological characteristics just by the measured resistivity value. Erroneous interpretations can be made when this fact is not taken into account.

This disadvantage does not prohibit successful lithological interpretations but in order to achieve good results prior information concerning the studied area should be considered. This prior information could be geologic or topographic maps of the area, or results from possible drilling, excavation results, or, in general, any kind of information that could enhance the knowledge of what may possibly be found beneath the soil. This information should be collected before the measuring

procedure to allow the optimum resistivity array and survey strategy to be chosen.

2.1.3 Formulation of the basic equations

The basic considerations that should be made in order to derive the equations that govern the flow of direct electrical current into the earth are the following:

- Ohm's Law

$$J = \sigma E \tag{2.2}$$

where:

J is the current density

σ is the conductivity

E is the electric field intensity

- The electric field E is the gradient of a scalar potential V.

$$E = -\nabla V \tag{2.3}$$

- The divergence of the current density J is zero when there are no charge sources or sinks in the medium, which is true for the earth's case.

$$\nabla \cdot J = 0 \tag{2.4}$$

From equations 2.2, 2.3

$$J = -\sigma \nabla V \tag{2.5}$$

So from Equations 2.5, 2.4 :

$$\nabla.(-\sigma \nabla V) = 0 \tag{2.6}$$

Which can be rewritten as [1]:

$$\nabla\sigma.\nabla V + \sigma \nabla^2 V = 0 \tag{2.7}$$

[1]From vector analysis it is known that for every scalar field z and field vector F, $\nabla \cdot (zF) = z \cdot / + \nabla z \cdot F$.

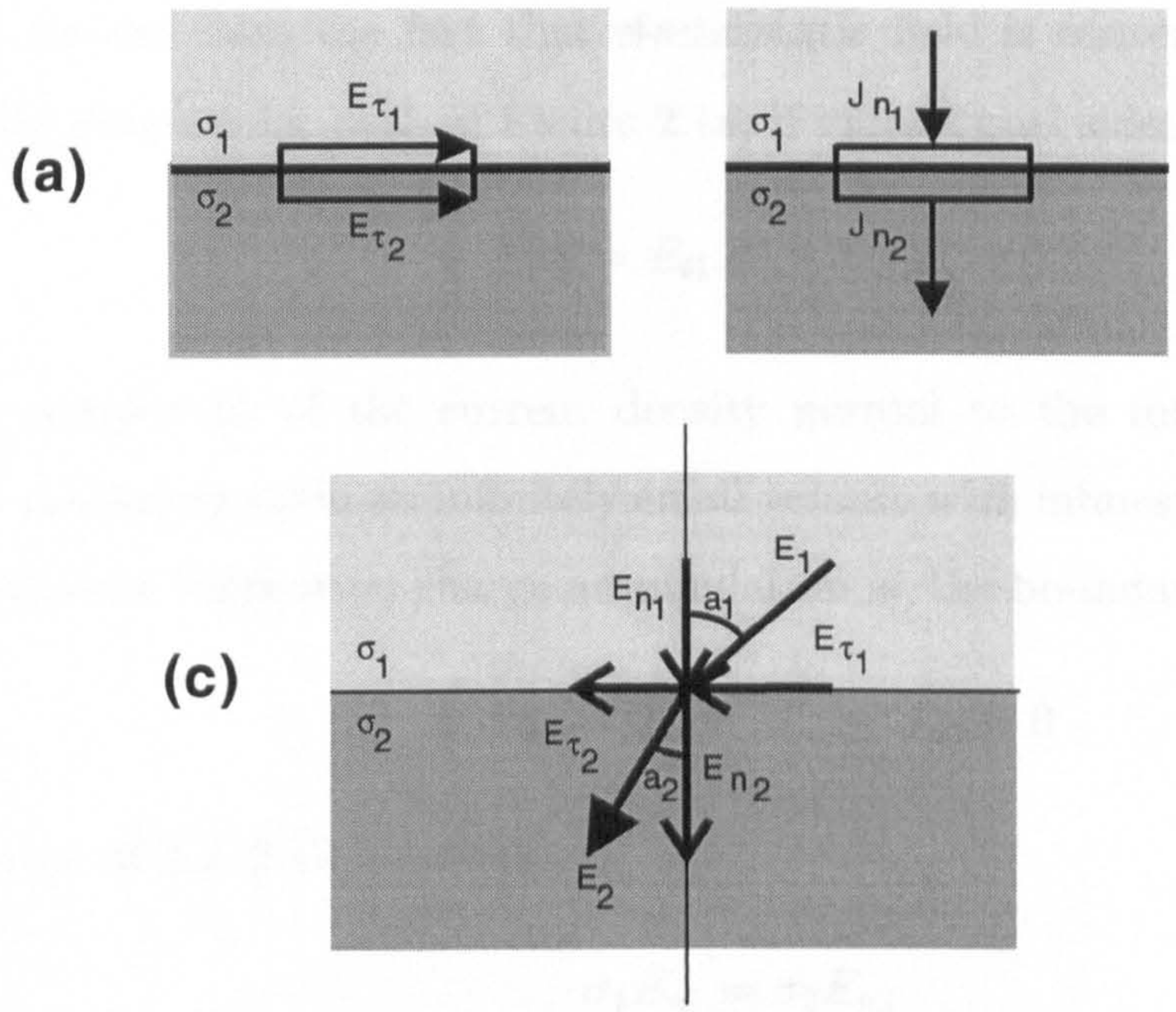


Figure 2.1: a) tangential electric field, b) normal current density, c) Snell's law for two media.

Equation 2.7 is a form of the well-known Poisson's equation, which describes the electrical flow in an inhomogeneous earth. Consider now the case of an homogeneous earth. The first part of the left side of 2.7 becomes zero since $\nabla\sigma$ is now zero

$$\nabla^2 V = 0 \tag{2.8}$$

This is Laplace's equation, and it is valid only for a homogeneous earth. Laplace's equation for spherical coordinates (r, θ, ψ) is :

$$\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial\psi^2} = 0 \tag{2.9}$$

The boundary conditions that hold at any contact between two regions of different conductivity σ_1, σ_2 are :

- The electric field tangential to the interface of the regions is continuous. In other words :

$$E_{t1} = E_{t2} \tag{2.10}$$

This derives from the fact that electrostatic field is conservative. Therefore for the rectangular path of Figure 2.1a, if infinestimal sides are assumed:

$$\oint E.dl = E_{t1}.dl - E_{t2}.dl = 0 \tag{2.11}$$

- The component of the current density normal to the interface is continuous. Assuming again an infinitely small volume with infinestimal sides (Figure 2.1b), since there is no charge accumulation at the boundaries :

$$\oint J.ds = 0 \quad or \quad J_{n1} - J_{n2} = 0 \tag{2.12}$$

Because of 2.2, 2.12 becomes:

$$\sigma_1 E_{n1} = \sigma_2 E_{n2} \tag{2.13}$$

A direct conclusion of the boundary conditions mentioned is that if a_1 , a_2 are the angles which are formed between the electric field lines in media σ_1 , σ_2, and the normal direction (Figure 2.1c) then:

$$E_{n1} = E_{t1}/(\tan a_1) \quad and \quad E_{n2} = E_{t2}/(\tan a_2) \tag{2.14}$$

Because of equations 2.10, 2.13, 2.14 :

$$\frac{\sigma_1}{\sigma_2} = \frac{E_{n2}}{E_{n1}} = \frac{\tan a_1}{\tan a_2} \tag{2.15}$$

This is actually Snell's law, which is also valid for seismic and electromagnetic waves.

Two other important considerations concerning the electrostatic field are:

a) the potential V is considered to be zero at a great distance from the source.

b) the family of equipotential curves are always orthogonal to the family of current curves.

Despite the fact that Poisson's equation is the one that describes the real situation in the resistivity method, since the earth is not homogeneous, Laplace's equation is extensively used. This is due to its simplicity, enabling the easy derivation of useful conclusions concerning current flow within the earth.

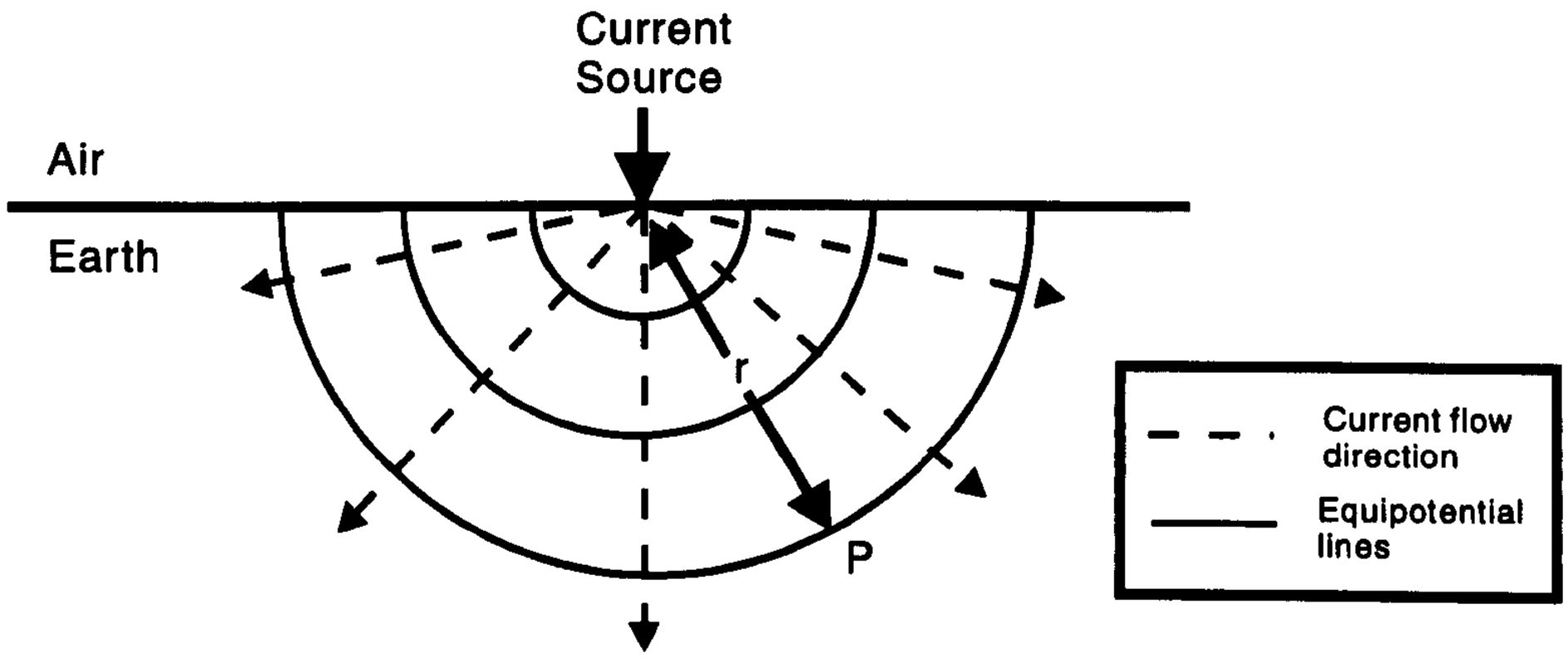


Figure 2.2: The equipotential surfaces and the direction of the current for one point source electrode.

2.1.4 Current flow in homogenous earth due to electrodes at the surface

Consider a single point source of current on the surface of a homogeneous earth - a semi-infinite space with conductivity σ. Laplace's equation for polar coordinates can be used in order to find the potential at every point P of the space with respect to its distance r from the source point. Because of the complete symmetry of the current flow with respect to θ and ψ directions, it can be assumed that the derivatives of these may be eliminated, so that 2.9 reduces to (Keller and Frischknecht, 1966):

$$\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right) = 0 \tag{2.16}$$

By integrating:

$$r^2\frac{\partial V}{\partial r} = C \tag{2.17}$$

or

$$\frac{\partial V}{\partial r} = \frac{C}{r^2} \tag{2.18}$$

By further integration of 2.18 :

$$V = \int \frac{dV}{dr}\,dr = \int \frac{C}{r^2}\,dr = -\frac{C}{r} + D \tag{2.19}$$

When $r \to \infty$ then $V \to 0$, hence D must be zero. The medium is homogeneous so that the current flows radially, crossing hemispherical equipotential surfaces. The total current density, J, of current with intensity I crossing the hemisphere surface, S, of radius r is given by J=I/S. Hence:

$$J = \sigma E = \frac{I}{2\pi r^2} \tag{2.20}$$

since $E = -\partial V/\partial r$

$$\frac{\partial V}{\partial r} = -\frac{I}{2\pi\sigma r^2} \tag{2.21}$$

From 2.18, 2.21 the value of C can be found:

$$C = \frac{I}{2\pi\sigma} \tag{2.22}$$

By substituting C in equation 2.19 and replacing conductivity, σ, by resistivity, ρ [$\sigma = 1/\rho$], the potential at every point of the homogeneous space is given by $J = I/S$. Hence:

$$V = \frac{I\rho}{2\pi r} \tag{2.23}$$

The equipotential lines for a point source electrode in a homogeneous space are shown in Figure 2.2. Consider now the case of two source point electrodes (this sort of arrangement is typically used in the resistivity method, since in effect a single electrode cannot carry current into the ground) on the surface of a semi-infinite space which has resistivity ρ. One of the electrodes is positive (A) and sends current I (source) into the ground while the other is negative (B) and collects the returning current $-I$ (sink). The potential measured at a point P, which is at a distance r_A and r_B from A and B respectively, is $V_P = V_{PA} + V_{PB}$ since potential is a scalar function. Taking into account 2.23:

$$V_P = \frac{I\rho}{2\pi r_A} + \frac{-I\rho}{2\pi r_B} \tag{2.24}$$

or

$$V_P = \frac{I\rho}{2\pi}\left(\frac{1}{r_A} - \frac{1}{r_B}\right) \tag{2.25}$$

In Figure 2.3 the equipotential curves and the current lines for a two point source configuration are presented.

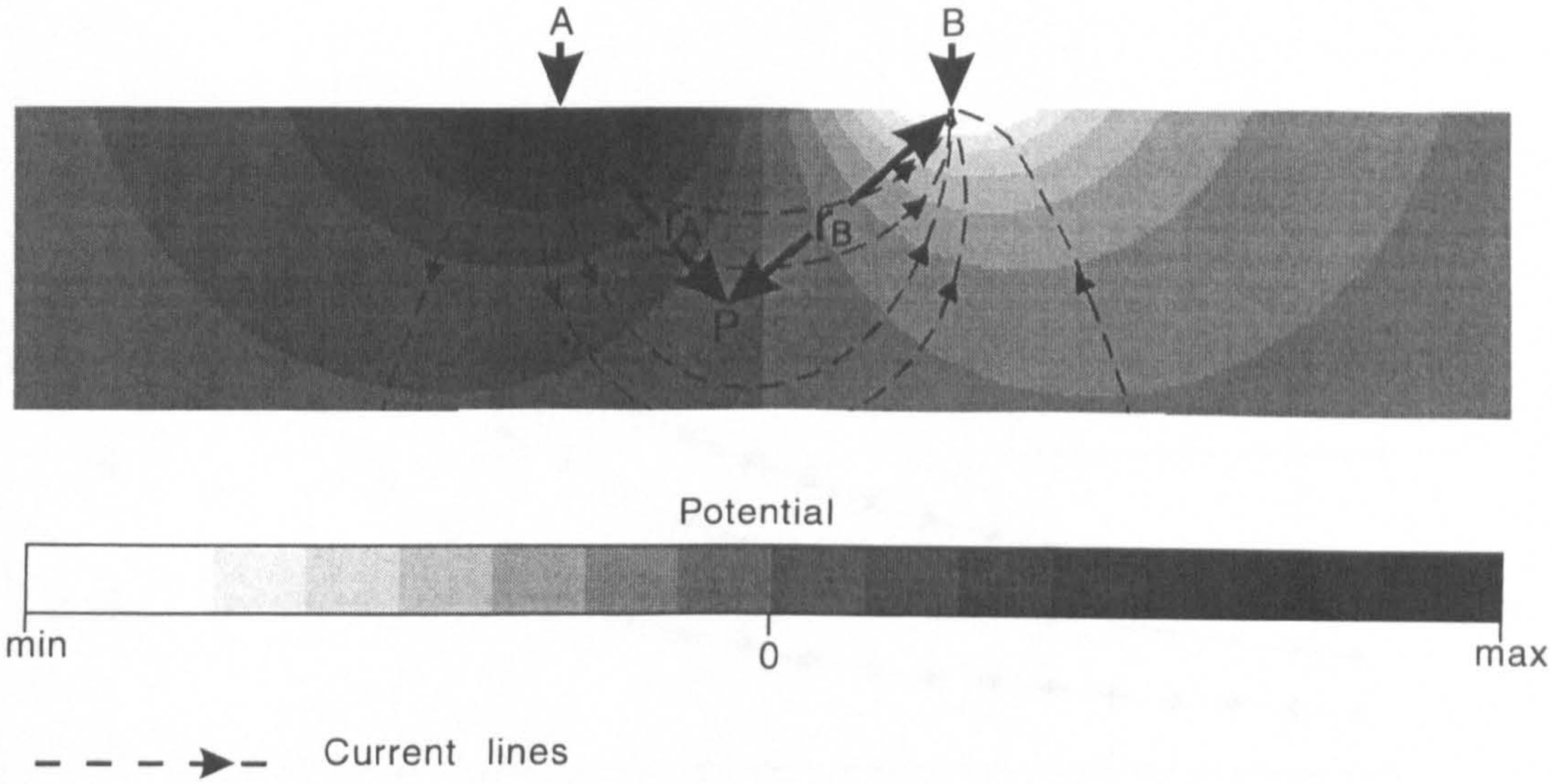


Figure 2.3: The equipotential surfaces and the current lines for two point source electrodes.

It is interesting to obtain an estimate of the depth that the current penetrates within an homogeneous half-space. This is of limited practical value because of two reasons : a) the investigation depth is also dependent on the positions of the potential electrodes (Roy and Apparao, 1971). b) The current penetration depends on the earth's inhomogeneity a very conductive top layer, for example, can restrict the current's depth penetration. However, it is still useful to get an indication of how the penetration depth varies with the separation of the sources.

Consider the case of two point electrodes A,B. If d is the distance between the sources, it can be shown that the fraction of current that penetrates below a depth z is (Telford et al., 1990):

$$\frac{I_x}{I} = (1 - \frac{2}{\pi}\, tan^{-1}(2z/d)) \tag{2.26}$$

By plotting the fraction of current that penetrates below a range of depths for different current probe separations (see Figure 2.4) it is obvious that the penetration depth increases as the distance between the probes gets bigger. Also from Figure

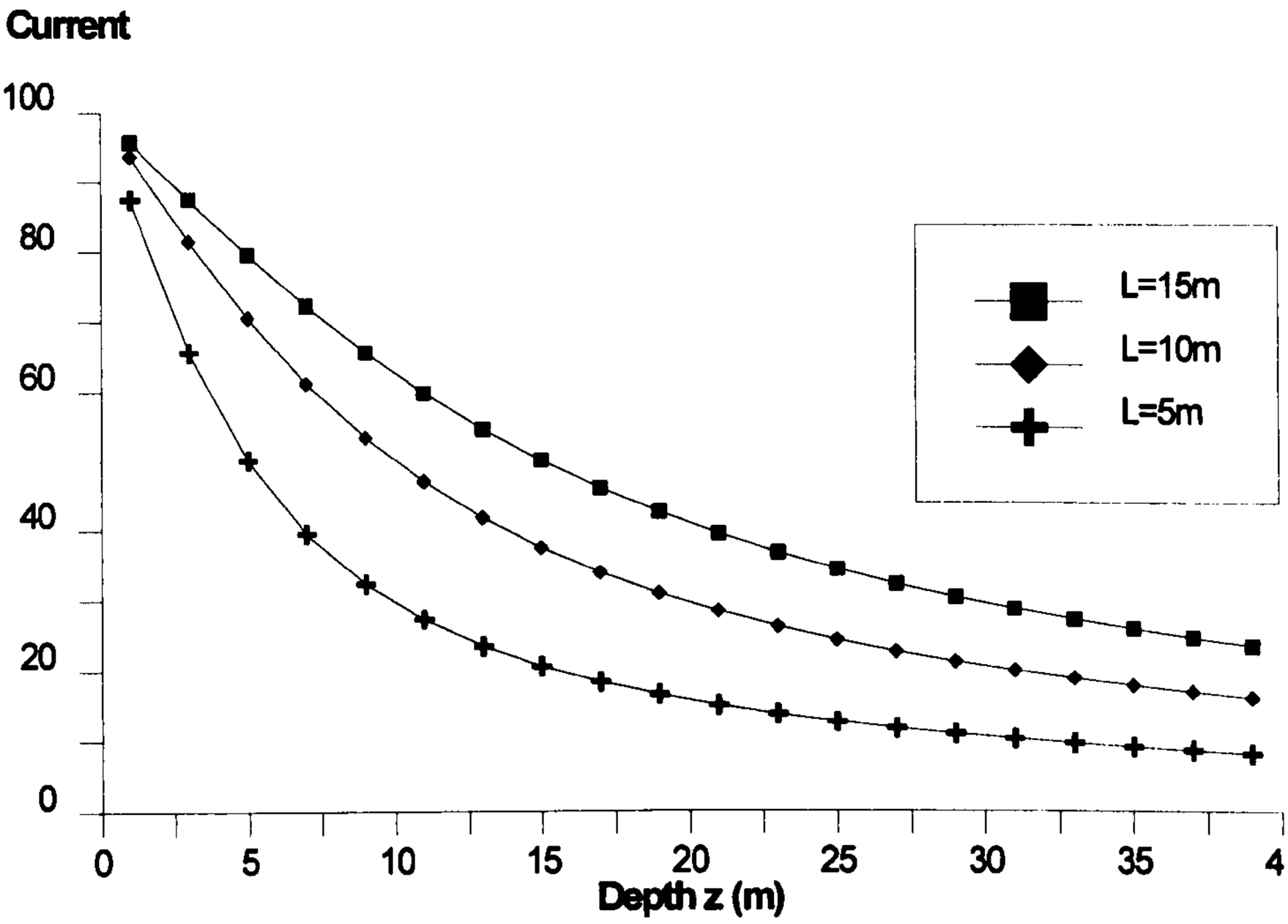


Figure 2.4: The % of current that penetrates below various depths for three different probe separations.

2.5 it can be concluded that 50% of the current penetrates below z=d/2 and just 30% below z=d.

These observations describe one of the most important features of the resistivity method: the effective penetration depth can be influenced by adjusting the separation of the current probes. This is a big advantage when the range of target depth is known (which is usually the case) but it can prove to be a limitation, either when no prior information is available or when the depth needed is extremely large.

2.1.5 Effect of anisotropy

So far the medium has been considered to be isotropic: in other words it is assumed that conductivity is the same in every direction. In reality, earth is anisotropic and therefore it appears to possess different conductivities depending on whether the current flows in one direction or another. The fact that conductivity is actually a

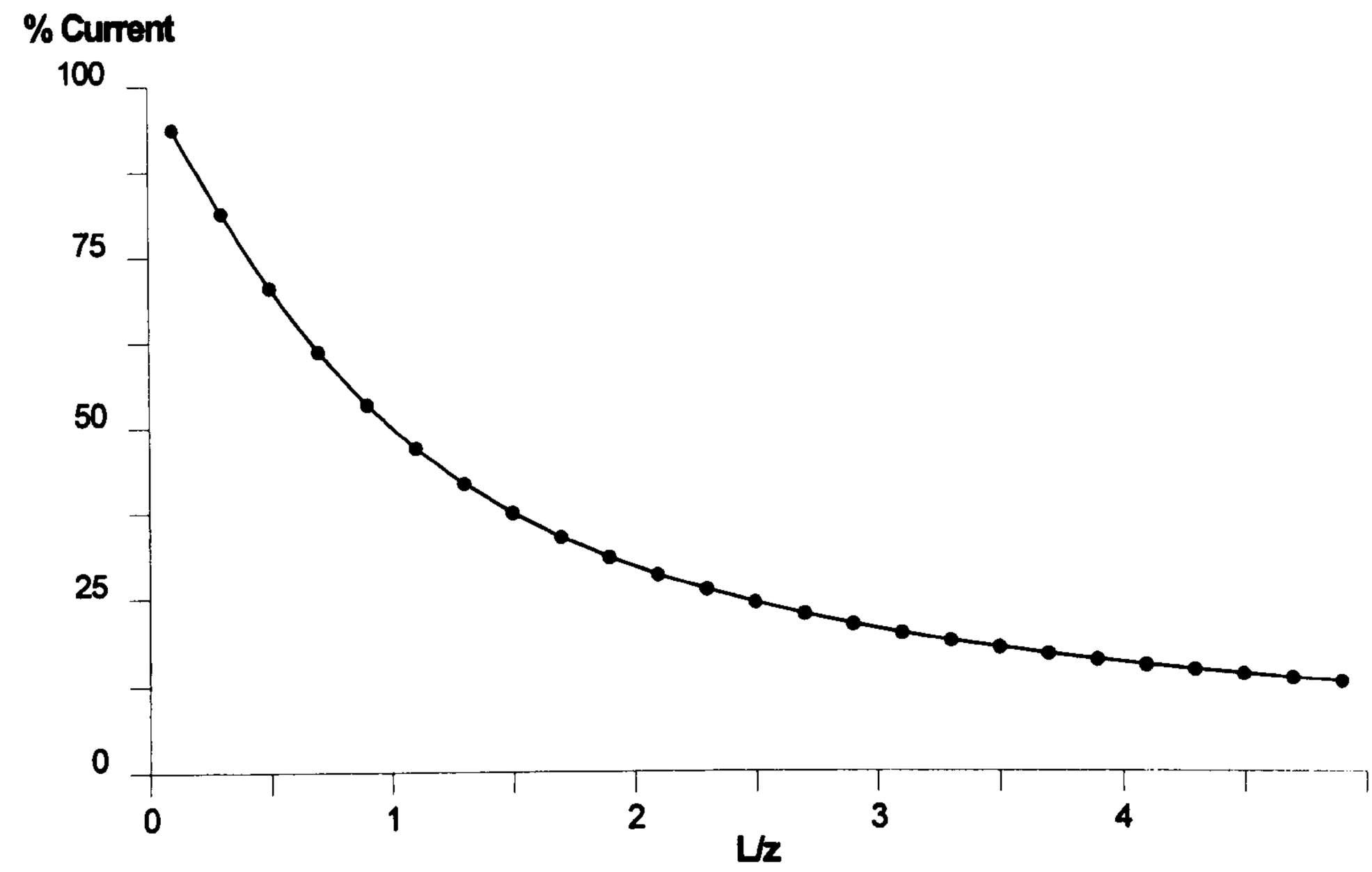


Figure 2.5: The % of current that penetrates below 15m while the probe separation increases.

second rank tensor and not a scalar, as was assumed so far, renders computations rather complicated. Therefore, only in rare cases is anisotropy taken into account in the mathematical formulation of the problem.

Fortunately, most of earth materials are not strongly anisotropic, and therefore the assumption of isotropy is not causing a big error. However, there are materials that do have highly anisotropical behaviour and so, even if it is difficult to include this factor in the computations, it should be taken into account while interpreting the data. Materials that display high anisotropy are either the stratified rocks (metamorphic rocks such as schists, shales, slates) or the highly fractured ones (e.g. limestone). As far as the stratified materials are concerned, the measured resistivity is smaller than the vertical resistivity and higher than the horizontal.

2.1.6 Apparent resistivity

In order to insert current into the ground two current electrodes (a source and a sink) have to be used. Those probes cannot be used at the same time to measure the drop of potential due to the variations in contact resistances between the current probes and the soil, so two different probes have to be used for that purpose. Therefore, in the resistivity method two current and two potential probes are used [2].

Assume a four terminal situation with current probes A,B and potential ones M,N over a homogeneous isotropic earth with resistivity ρ (see Figure 2.6a). The potential at M and N due to A,B can be found by using equation 2.25 and it is :

$$V_M = \frac{I\rho}{2\pi}\left(\frac{1}{AM} - \frac{1}{BM}\right) \tag{2.27}$$

$$V_N = \frac{I\rho}{2\pi}\left(\frac{1}{AN} - \frac{1}{BN}\right) \tag{2.28}$$

Therefore, the drop of potential ΔV will be:

$$\Delta V = V_M - V_N = \frac{I\rho}{2\pi}\left(\frac{1}{AM} - \frac{1}{BM} - \frac{1}{AN} + \frac{1}{BN}\right) = \frac{I\rho}{2\pi}\ G \tag{2.29}$$

Solving 2.29 for ρ:

$$\rho = 2\pi\ \frac{\Delta V}{I}\ \frac{1}{G}\ =\ 2\pi\ \frac{R}{G} \tag{2.30}$$

Where:

- ΔV = the potential drop
- I = the intensity of the inserted current
- R = the measured resistance ($R = \Delta V/I$)
- ρ = the resistivity of the homogeneous ground
- $AM,\ AN,\ BM,\ BN$ = the distances between the current and potential probes

[2] the following notation (which is quite common in literature) will be used when referring to the probe arrays : I^+=A, I^-=B, V^+=M, V^-=N.

- G = the so-called geometrical factor

In the case of a homogenous ground, for any possible four probe arrangement the geometrical factor, when applied to the measured resistance (R), yields the true resistivity.

Reciprocity theorem in resistivity

It can be seen from equation 2.29 that if the current electrodes A,B are interchanged with the potential M,N ones there will be no change of G. Hence, for a homogeneous earth, the potential difference ΔV will remain unchanged. According to Parasnis (1990) this is a special case of the Helmholtz'z reciprocity theorem and is also valid for an inhomogeneous earth by assuming that it consists of any number of separate homogeneous isotropic regions. Another condition that should be met is that the current passing through the new current electrodes should be kept the same. Most of the time reciprocity is verified in practice, however in noisy environments (geological noise) and for certain type of arrays (Park and Van, 1991) the reciprocity concept can be violated. In cases where real data are being interpreted by algorithms which make use of the reciprocity principle it is essential to take into account the degree in which the data satisfy the reciprocity test. In other words, the reciprocity test is indicative of the quality of the obtained data.

Apparent resistivity for inhomogeneous ground

There are many possible resistance measurements that can be obtained by any four-probe array, whether linear or not. However, as a direct consequence of the reciprocity theorem, it can be shown that only three of them can possibly be independent. The three basic electrode arrangements which provide the R_α, R_β and R_γ resistances are labelled by α, β, and γ accordingly [3] and are shown in Figure 2.6b.

[3]The respective resistivities are ρ^α, ρ^β and ρ^γ.

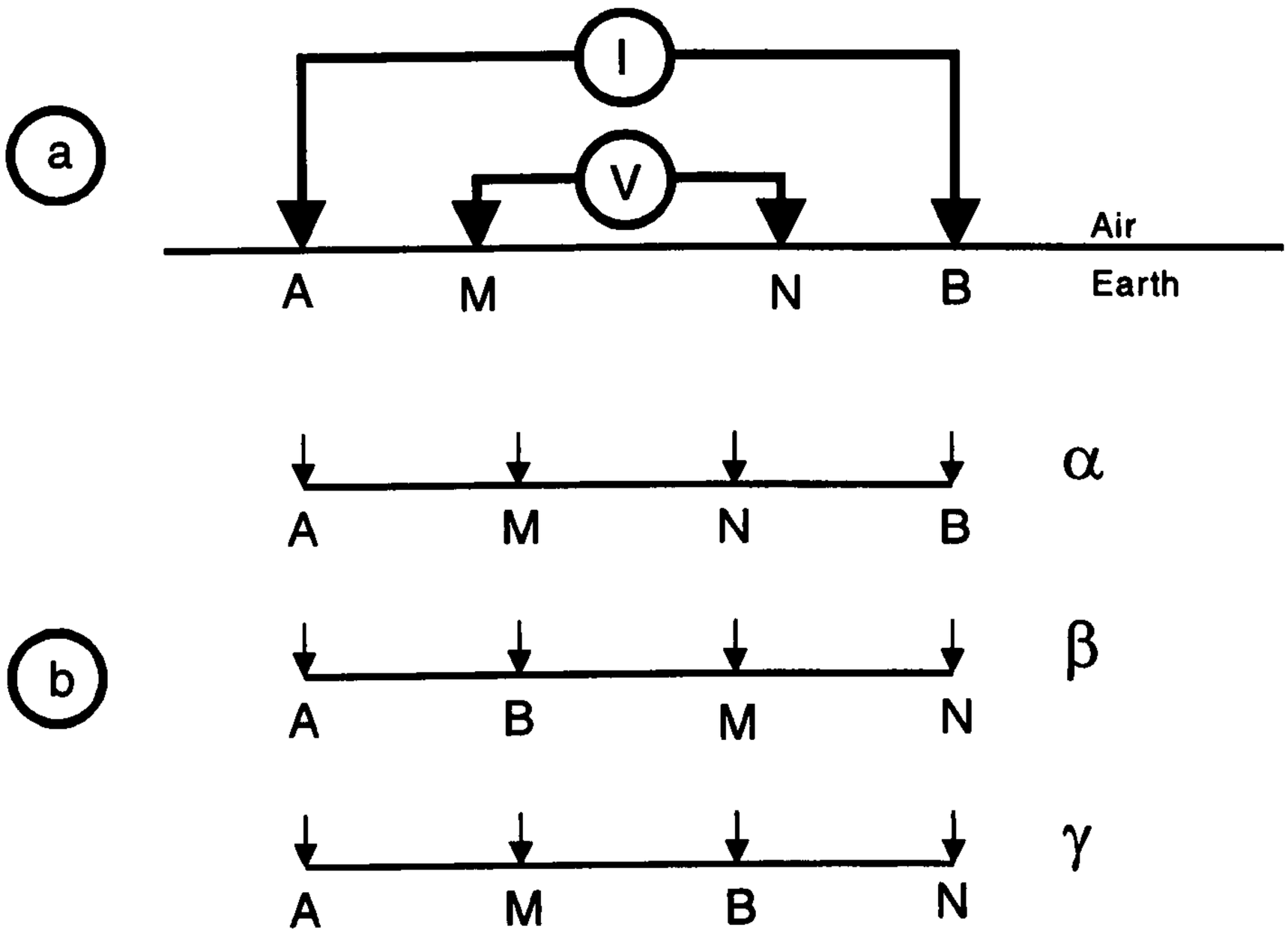


Figure 2.6: a) A 4 electrode array , b) The α, β and γ configurations.

By superposition, it can easily be shown that the following relation holds for these three resistances:

$$R_\alpha - R_\beta = R_\gamma \tag{2.31}$$

Consider now the case of an inhomogeneous ground. Equation 2.30 no longer represents the true resistivity of the ground. The value that it gives in the case of an inhomogeneous ground is called apparent resistivity (AR), ρ_a. Apparent resistivity is sometimes described as representing the resistivity that the ground would have had if it were, in fact, homogeneous.

It is interesting at that point to discuss the physical meaning of the AR. Clark (1990) characterizes AR as the average resistivity of the volume of the ground that is being affected by the inserted current, Robinson and Coruh (1988) states that AR is a weighted average of the true resistivity of the subsurface. This is by no means mathematically true (Telford et al, 1990), however, in a broad sense this concept can be useful in the interpretation of many simple problems. It should be born in mind though, that AR is an artificial concept and there are situations where regarding it

as the weighted mean of the true resistivities can be totally unrealistic.

Such an example (Carpenter and Habberjam, 1956) is shown in Figure 2.7: The four probe arrangement shown at the top of the figure was used in the three possible modes to take resistivity measurements over a two layered earth with the bottom layer 100 times more resistive than the top. The AR for the α, β and γ configurations were calculated for different depths (z) of the first layer using the technique of electrical images [4]. From this example it can be clearly seen that the AR for the γ configuration (ρ_a^γ) can obtain negative values. Any connection of these measured AR values with the real resistivity (which is by definition always positive) is unrealistic. This example, though somewhat extreme, indicates the artificiality of the AR concept and the fact that interpretations based on the AR should be made with caution.

2.2 Resistivity Arrays and Instrumentation

In this section the most common resistivity arrays are presented. A general description of the resistivity measuring modes will be also given and a comparison of the resistivity arrays is presented. The resistivity instrumentation and the automated measuring systems will be described and, finally, problems encountered in resistivity surveying will be mentioned.

2.2.1 Resistivity arrays

The way in which the current and potential electrodes are arranged on the earth's surface is called a resistivity array. A large number of probe arrangements has been suggested, but only a few are used extensively. This happens because, as will

[4]the technique of electrical images (Telford et al., 1990) is based on the analogy between the electrical situation and optics, and is a quite uncomplicated and effective method for solving a limited number of simple potential problems.

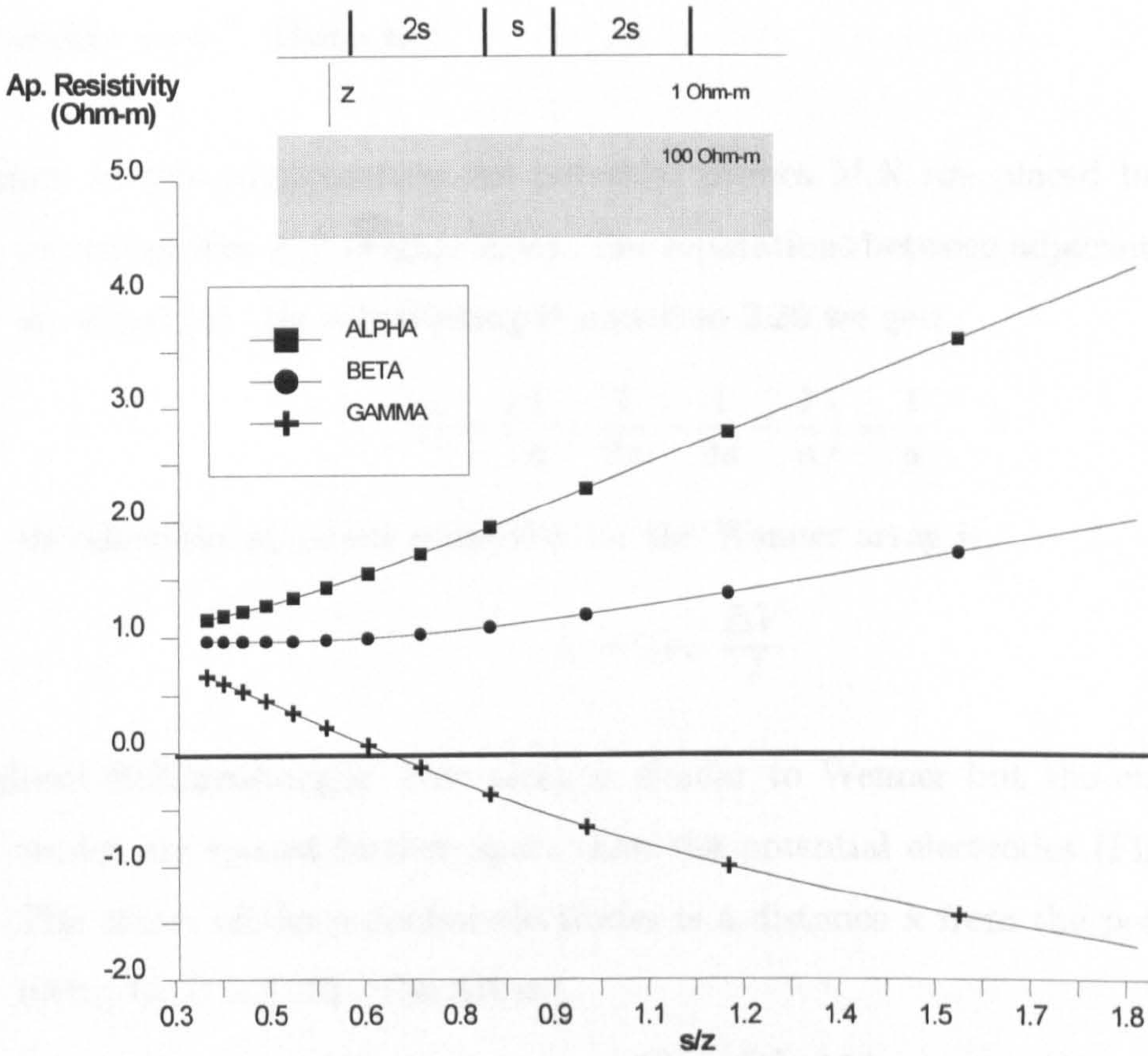


Figure 2.7: The apparent resistivity curves for the α, β and γ configurations measured over a 2-layered earth model (top) for variable thicknesses of the first layer.

shown later, the use of an array is based both on theoretical and practical criteria, and therefore arrays that do have theoretical merits are often not used because of their practical drawbacks. For instance, the penta-electrode arrangement in theory performs better than many conventional arrays (Bernabini et al., 1987) but is hardly ever used since its application is not practical.

The main characteristic of an array is its geometrical factor, which is uniquely related to the relevant distances between the probes. The majority of the popular arrays involve electrodes that are in line and have distances governed by internal symmetry since, in this way both positioning and data interpretation become much easier.

In this work, the presentation of the arrays will be restricted to these which are

most widely used [5]. These are:

Wenner In this configuration the potential probes M,N are placed between the current probes A,B (Figure 2.8a). The separations between adjacent electrodes are equal (a). By substituting in equation 2.29 we get:

$$G = \left(\frac{1}{a} - \frac{1}{2a} - \frac{1}{2a} + \frac{1}{a}\right) = \frac{1}{a} \tag{2.32}$$

therefore the apparent resistivity for the Wenner array is:

$$\rho_a = 2\pi a \, \frac{\Delta V}{I} \tag{2.33}$$

Gradient Schlumberger This array is similar to Wenner but the current electrodes are spaced further apart than the potential electrodes (Figure 2.8b). The centre of the potential electrodes is a distance x from the point of symmetry O. If L$\geq$ 10 l the AR is:

$$\rho_a = \frac{\pi}{2l} \, \frac{(L^2 - x^2)^2}{(L^2 + x^2)} \, \frac{\Delta V}{I} \tag{2.34}$$

Schlumberger This is similar to the previous array but all electrodes are placed symmetrically around a point O (Figure 2.8b). If the distance between the current electrodes is 2L and the distance between the potential probes is 2l with L$\geq$ 10 l then the AR is:

$$\rho_a = \frac{\pi L^2}{2l} \, \frac{\Delta V}{I} \tag{2.35}$$

Dipole-Dipole In this array (Figure 2.8c) the current electrodes (current dipole) are separated from the potential ones (potential dipole). Both dipoles have stable separations (a) and the distance between them is na (proportional to a). The AR is:

$$\rho_a = -\pi n(n+1)(n+2)a \, \frac{\Delta V}{I} \tag{2.36}$$

[5]In the literature there is no general agreement on the naming of the various arrays, so for each array the name with which it is most widely known will be used.

Pole-dipole The potential probes are between the current probes but one of the current probes, say B, is located a great distance from the remaining three (Figure 2.8d). In this way the distances BM, BN are considered to be infinite and therefore the terms 1/BM, 1/BN in the geometrical factor G are zero. If a is the AM distance and b the AN distance the AR is :

$$\rho_a = \frac{2\pi ab}{b-a}\,\frac{\Delta V}{I} \tag{2.37}$$

In the case where the distance b is proportional to the distance a ($b = na$) equation 2.37 becomes:

$$\rho_a = 2\pi n(n+1)a\,\frac{\Delta V}{I} \tag{2.38}$$

Pole-pole This is a further variation of the pole-dipole array and is obtained by also moving one of the potential electrodes, say N, a great distance from the remaining A, M probes (Figure 2.8e). Thus the distances that are considered to be infinite are now BM, BN, AN. If AM=a, the geometrical factor G becomes G=1/a, which is the same as for the Wenner array, so the AR is

$$\rho_a = 2\pi a\,\frac{\Delta V}{I} \tag{2.39}$$

A further modification of the pole-pole array known as **twin-probe** is obtained by placing the remote probes (B, N) themselves close together (Figure 2.8f). The distance BN is not considered infinite any more and if BN=b the geometrical factor becomes

$$G = \frac{a+b}{ab} \tag{2.40}$$

and the AR

$$\rho_a = 2\pi\frac{ab}{a+b}\,\frac{\Delta V}{I} \tag{2.41}$$

In the case that BN=AM=a equation 2.41 becomes:

$$\rho_a = \pi a\,\frac{\Delta V}{I} \tag{2.42}$$

Square Array In this array, the four probes occupy the vertices of a square of side a. The three possible arrangements α, β and γ for the square array can be seen in Figure 2.8g. The apparent resistivities ρ_a^α, and ρ_a^β are expressed as

$$\rho_a^\alpha = \rho_a^\beta = \frac{2\pi\ a}{2\ -\ \sqrt{2}}\ \frac{\Delta V}{I} \tag{2.43}$$

The geometrical factor for the γ arrangement is not defined, so the AR for the square γ array is not defined as well. Since both square α and β configurations have identical geometrical factors and because of equation 2.31, for the case of a homogenous ground the Rγ resistance is zero.

The most interesting feature of the square array is its ability of the array to highlight directional variations of the subsurface resistivity. This orientational variation of the electrical properties of the subsurface can be expressed by the Azimuthal Inhomogeneity Ratio (AIR) which is given by

$$AIR = \frac{2\ R_\gamma}{R_\alpha\ +\ R_\beta} = \frac{2\ (R_\alpha\ -\ R_\beta)}{R_\alpha\ +\ R_\beta} \tag{2.44}$$

It is important at this point to discuss the assumption of the "infinite" distance. Both physically and practically there is no way of achieving an "infinite" distance. Therefore the term "infinite" is being used in a less strict sense, representing the fact that if a distance is far grater than another it can be omitted from the calculation of the geometrical factor without causing a significant error.

Taking, for example, the case of the pole-dipole array with A, M, N equally spaced so that AM=MN=a and B spaced so that BN=20a, the percentage error that will occur by omitting BN from the calculation of the geometrical factor can be calculated easily: $G_{real} = 0.498(1/a)$ while $G_{theor.} = 0.50(1/a)$, and hence the error is approximately 0.4%. It is obvious that the greater the distance the smaller the error, but since in surveys the length of the cables used is limited, there is a minimum distance that keeps errors within acceptable limits. For pole-dipole arrays this distance is 10a, while for the pole-pole it is 30a.

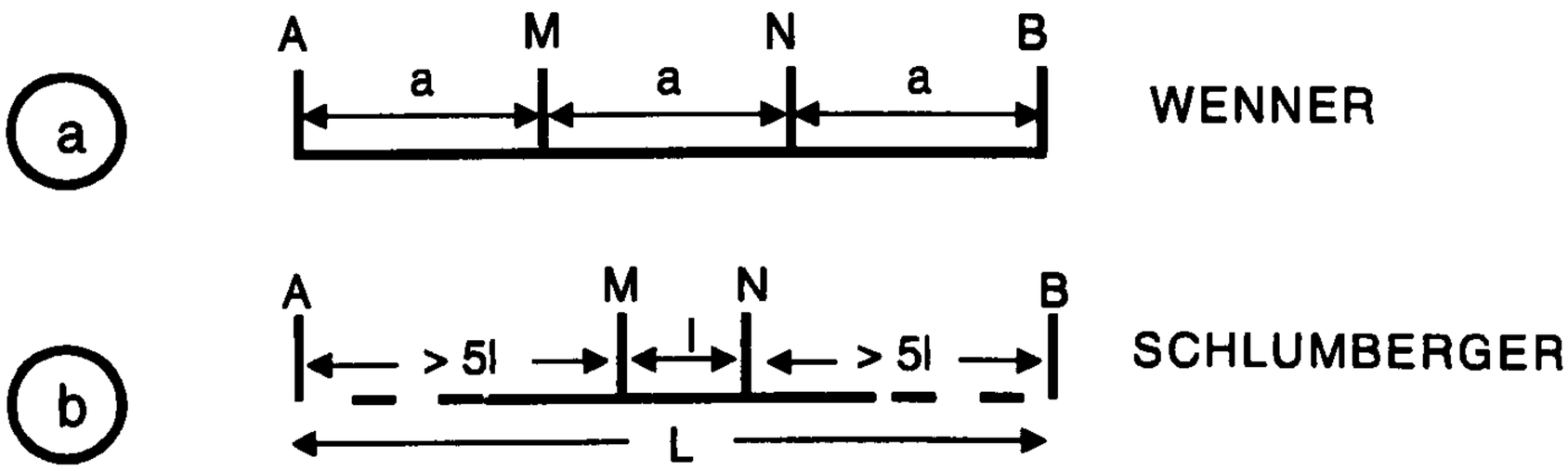


Figure 2.8: Common resistivity arrays : a) Wenner, b) Schlumberger, c) dipole-dipole, d) pole-dipole, e) pole-pole, f) twin-probe, g) square array.

2.2.2 Resistivity measuring modes

There are three measuring procedures in resistivity prospecting, the choice of which depends on whether one desires to locate variations of resistivity with depth, with lateral extent or with both depth and lateral extent. In the first case, the procedure is called vertical sounding, the second case lateral profiling, while the third case is a combination of sounding and profiling.

Vertical Sounding This procedure is based on the fact (section 2.1.4) that the effective depth of current penetration increases as the probe separation is increased. In this procedure arrays with internal symmetry are used. At every measurement the separation of the measuring electrodes is increased relative to a fixed centre O. In this way, a series of measurements is produced that "broadly" correspond to the variation of resistivity at various depths below the centre O.

Lateral Profiling In the lateral profiling (LP) procedure the separation of the probes remains stable throughout the survey while a series of measurements are taken by moving the array in a lateral direction. By plotting the resistivity values along the measured profile the anomaly caused by a subsurface feature can be picked out. It is clear that in this way only information about the lateral variation of resistivity can be collected since the current probe separation remains the same - no information about the vertical extent of the anomaly can be obtained.

Combined sounding and profiling This is a combination of both lateral profiling and vertical sounding procedures. In this way information about both vertical and lateral variations of resistivity can be acquired. The data acquisition pattern can be viewed either as a series of profiles over the same area with increasing probe separation or as a series of adjacent vertical soundings along the survey line. One of the main characteristics of this measuring mode

is that, compared to the other procedures, a larger amount of measured data (and therefore potentially useful information) is collected.

So far, the main resistivity measuring modes have been defined: all of these modes are traditionally applied in a linear direction (survey line). However, other measuring modes exist involving measurements acquired in a non-linear fashion (Bibby and Hohmann, 1993). These sort of measuring modes, although not widely used, can be quite effective for certain type of surveys (e.g. geothermal field exploration). Furthermore, measuring modes involving a borehole electrode (miss-*à*-la-mass method) can be employed as well. Recently, the development of automatic multiplexing allowed the use of a series of borehole probes. In this way borehole-to-surface or borehole-to-borehole data sets can be obtained (Zhao et al., 1986).

2.2.3 Evaluation of the resistivity arrays

The choice of a particular resistivity array for a survey is based upon considerations regarding theoretical merits and demerits of the array, while, at the same time practical aspects, such as the purpose of the survey, the survey cost, and the available equipment and personnel are equally important.

Ward (1989) presented a review on the evaluation of several resistivity arrays. He proposed fourteen possible criteria for selecting a resistivity array. He suggested, however, that this evaluation is limited since the published literature contained inadequate information for a truly objective comparison of the arrays. The most important of his criteria and the subsequent evaluation are shown in Table 2.3. It has to be noted that most of the evaluation results have been drawn from systematic comparisons of models calculated using numerical techniques [6] (Coggon, 1973; Dey et al, 1975).

[6] A detailed description of these techniques will be given at Chapter 3.

Array	S/N ratio	Lateral location	Resolution of hor. layers
Wenner	1	5	1
Schlumberger	2	4	1
Dipole-dipole	5	2	2
Pole-dipole	4	3	2
Gradient	3	1	5
code: 1=best, 5=worst			

Table 2.3: Resistivity array evaluation (after Ward, 1990).

Depth of investigation

A significant criterion for choosing an array is the investigation depth of an array: the approximate maximum depth of a body that will be detected by the array. Field geophysicists mainly use the empirical depth of an array. An estimate of this depth is based on practical experience acquired by comparing the measured data with drilling or excavation results. It is very difficult to give an absolute value of the investigation depth for two reasons:

a) The measured signal in resistivity prospecting is due to variable contributions from various depths, so there is no particular depth which contributes uniquely to the measured signal.

b) despite the fact that theoretical values concerning the depth of investigation can be calculated for a homogeneous earth, the actual exploration depth depends on (the unknown) characteristics of the existing subsurface features (inhomogeneities).

A theoretical approach to investigation depth is the concept of a "depth of investigation characteristic" (DIC) suggested by Roy and Apparao (1970). Despite the reasonable criticism (Parasnis, 1990) that the DIC is not objective since it applies only to a homogeneous earth, we will further discuss the concept for two reasons:

a) It can give a qualitative insight into the relative behaviour of the various resistivity arrays.

b) It has been used as the basis of approximate reconstruction methods which will be examined in a later chapter.

Roy and Apparao (1970) showed that the depth of investigation of resistivity arrays is not only dependent on the current probe separation, but it is also affected by the positions of the measuring probes. Therefore, instead of examining the current distribution within the ground, they calculated the contribution to each elementary volume of earth to the signal ($\Delta V/I$) observed at the surface by integrating this contribution over a thin horizontal layer parallel to the surface. Calculation of this contribution for several depths produced the (DIC) curve which for a generalized four electrode array has the form (z is the depth):

$$\begin{aligned} DIC &= \int 4z\,\rho I\Big[\frac{1}{(AM^2+4z^2)^{3/2}} - \frac{1}{((MN+NB)^2+4z^2)^{3/2}} \\ &- \frac{1}{((AM+MN)^2+4z^2)^{3/2}} + \frac{1}{(NB^2+4z^2)^{3/2}}\Big]dz \end{aligned} \tag{2.45}$$

The DIC curves for Wenner, dipole-dipole and pole-pole arrays are shown in Figure 2.9. DIC begins with zero value for zero depth, reaches a maximum and finally reaches zero at large depths. The depth of investigation, according to Roy and Apparao (1970), is the depth where the DIC curve attains its maximum - the depth that gives the largest contribution to the signal received at the ground surface.

Further, the DIC curve gives an indication of the resolution with depth: in general, the bigger the width of the curve the smaller the resolution. The depth resolution is defined as the inverse of the width of the DIC curve at its half-maximum points. The investigation depth and the resolution relative to the distance L between the two external active (not at "infinite" distance) probes of the array are presented in Table 2.4.

Edwards (1977) compared these results with empirical ones and suggested that the most representative choice for the investigation depth of an array is not the one given by the maximum of the DIC curve (maximum depth) but the one given by the

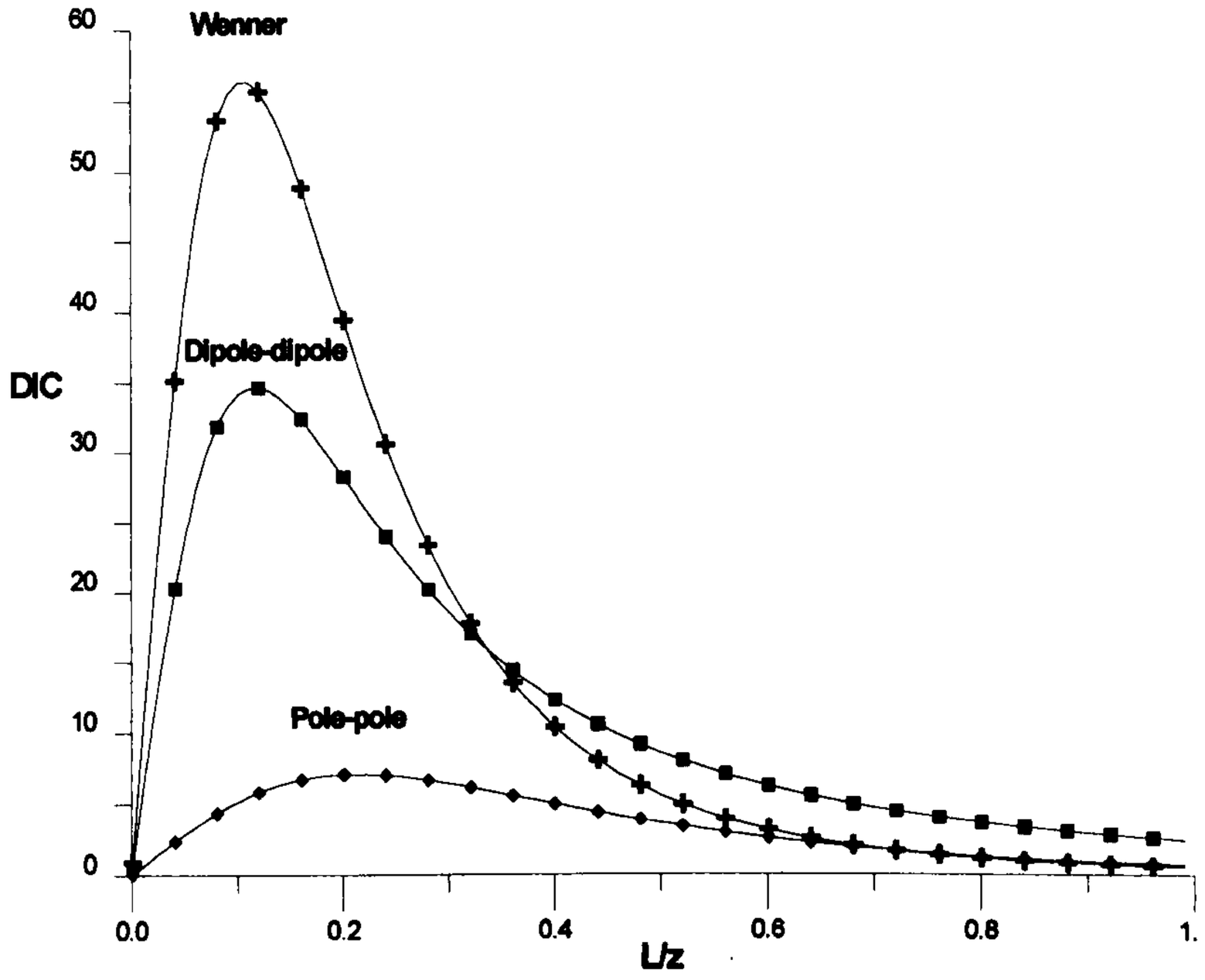


Figure 2.9: DIC curve for the Wenner, dipole-dipole and pole-pole arrays.

point of the curve which corresponds to the depth that one-half of the total signal originates - this depth is the "median" depth. The results of Edwards (see Table 2.5), which were also supported by a later work of Barker (1989), are somewhat more optimistic than those of Roy.

No matter which results are chosen, some interesting remarks can be made about the characteristics of the arrays:

a) The arrays which use remote probes (twin-probe, pole-dipole) have by far the largest depth of investigation. This advantage does not mean that they are superior to the other arrays since this feature is counterbalanced by their lack of resolution with depth. In general, this characteristic renders them ideal for reconnaissance investigation but prohibits their use in high resolution vertical investigations.

b) Wenner, Schlumberger and Dipole-Dipole arrays appear to have high a vertical resolution and an adequate depth of investigation and hence they can be used whenever high resolution with depth is needed.

c) Gradient Schlumberger was proved to have a very large investigation depth and therefore a very low resolution with depth.

Array	Depth	Resolution
Wenner	0.11 L	1/2.25
Schlumberger	0.125L	1/2.45
Dipole-dipole	0.195L	1/3.1
Twin-probe	0.35 L	1/8.4

Table 2.4: Depth of investigation and resolution of some resistivity arrays (after Roy and Apparao, 1970; Roy, 1971).

Array	Depth
Wenner	0.17 L
Schlumberger	0.195L
Dipole-dipole	0.25L
Twin-probe	0.77L
Pole-dipole	0.52L

Table 2.5: Depth of investigation of some common resistivity arrays (after Edwards, 1977).

The method for calculating an investigation depth can be extended to simple inhomogeneous models (Apparao et al., 1992), but these results still cannot be used to derive general conclusions.

2.2.4 Resistivity instrumentation

In order to measure the resistivity of the earth special instruments called resistivity meters (RM) are used. Further, equipment such as cables and probes are also needed. Two current probes are used to insert current into the ground and due to the large contact resistance between the probes and the ground, separate voltage probes should be used: four probes in all. The resistivity meter then measures the resistance R which is simply the ratio of the voltage across the potential probes V_{mes} to the intensity I of the inserted current. Hence the apparent resistivity can be

readily found (see equation 2.30). The resistivity meter has two main components a) The power source and b) The measuring system (see Figure 2.10a).

Power source

For obtaining a resistivity measurement, the test current has to pass through a circuit including the earth resistance and the resistances of the current probes. The total resistance could be several hundred ohms and therefore the voltage used should be fairly high. In actual instruments it is between 50 and 500 volts. The power source could be either a DC or an AC one.

DC source: If a direct current source is used, two problems have to be faced:

a) Polarization effect: There is a concentration of anions around the negative electrode and cations around the positive one. This build-up of ions will gradually increase with time and therefore the measured resistance between the probes will gradually increase too.

b) Spontaneous potential: Natural electrical currents of electrochemical origin exist in the subsurface. These currents will increase the value of the measured potential.

The problem of the polarization effect is tackled by reversing periodically the DC polarity or by using special porous porcelain probes. The spontaneous potential effect should be measured before the source is turned on and then subtracted from the measured voltage when current is inserted. Modern instruments do this procedure automatically.

AC source: When an AC current is used the accumulation of ions does not occur and so electric polarization is eliminated. Furthermore, the AC signal nullifies the effects of spontaneous potential (at every current reversal the spontaneous potential current increases or decreases the measured potential by equal amounts). The frequencies used in the AC source must be low enough for effects such as electro-

magnetic induction to be minimal, and the selection of the frequency is also related to the required depth of the investigation (the highest the frequency the lower the depth of investigation). Finally, when an AC source is used, capacitance effects and EM coupling between current and potential probes also have to be considered.

Measuring systems

Theoretically, a simple ammeter and voltameter are adequate for measuring the earth's resistivity. However, quite advanced instruments have been developed. The design of the measuring system is closely related to the purposes of the instrument: RMs used in archaeology have totally different specifications from those used in geological applications. In the following table some typical specifications of RM used in archaeological and geological surveys are shown [source: Geoscan, Geonics, OYO].

SPECIFICATIONS	RM (archaeol.)	RM (geol.)
OUTPUT CURRENT	0.1-10 mA	1-200 mA
OUTPUT VOLTAGE	50 V max.	200 V max.
FREQUENCY	0.1-150 Hz	DC/ 100 Hz max.
POTENTIAL RANGE	20-200-2000 mV	20-200-2000 mV

Most of the modern automatic instruments make use of the constant current principle. The current source is designed to maintain a constant current in to the ground by varying its voltage in response to the changes in the resistance. It should be noted that these changes may be of a wide range (e.g. 0.05-2000 Ohms) As long as the potential difference reading is taken the resistance can be readily found. Modern instruments have also other characteristics such as measurement recording systems, high accuracy via averaging, measurement noise reduction devices etc.

Automatic RM

In recent years the development of the RMs has been oriented towards systems which automate the resistivity measuring procedure. This is being achieved by the use of automatic multiplexing. Automatically multiplexed RM have main measuring units identical to that of the conventional RM. This measuring unit is connected to a switching unit which contains computer controlled relay cards which are used to connect the RM with any four of a series of electrodes via a multi-core cable. The entire measuring procedure is controlled by a computer which not only "decides" which probes will be connected but also can be used to trigger the RM and to store the measurements, as well as to change the RM's settings (e.g. current intensity, measurement sampling etc.).

There are two main ways with which the multiplexing can be achieved:

a) Central switching unit. In this approach the channels in the multi-core cable should be as many as the probes used for the survey. This increases the cable cost and limits the total number of the probes that can be used since the number of channels in a multi-core cable is fixed.

b) Distributed switching units. This switching mode involves computer controlled switching circuits at every probe point [7]. In this way multi-core cables with a limited and stable number of channels can be used (e.g. five channels can be enough: two for the current transmission, two for the potential measurements and one for the computer signal) and therefore the maximum number of probes that can be used is theoretically unlimited.

From a theoretical point of view distributed switching units are superior to central switching units as far as cabling cost and flexibility is concerned, however the total cost of the distributed switching unit itself can be higher than that of a central one. In practice, central switching units are less sensitive to rough field conditions

[7]For practical reasons these switching circuits are mounted on the probes this sort of probes are known as "smart" probes.

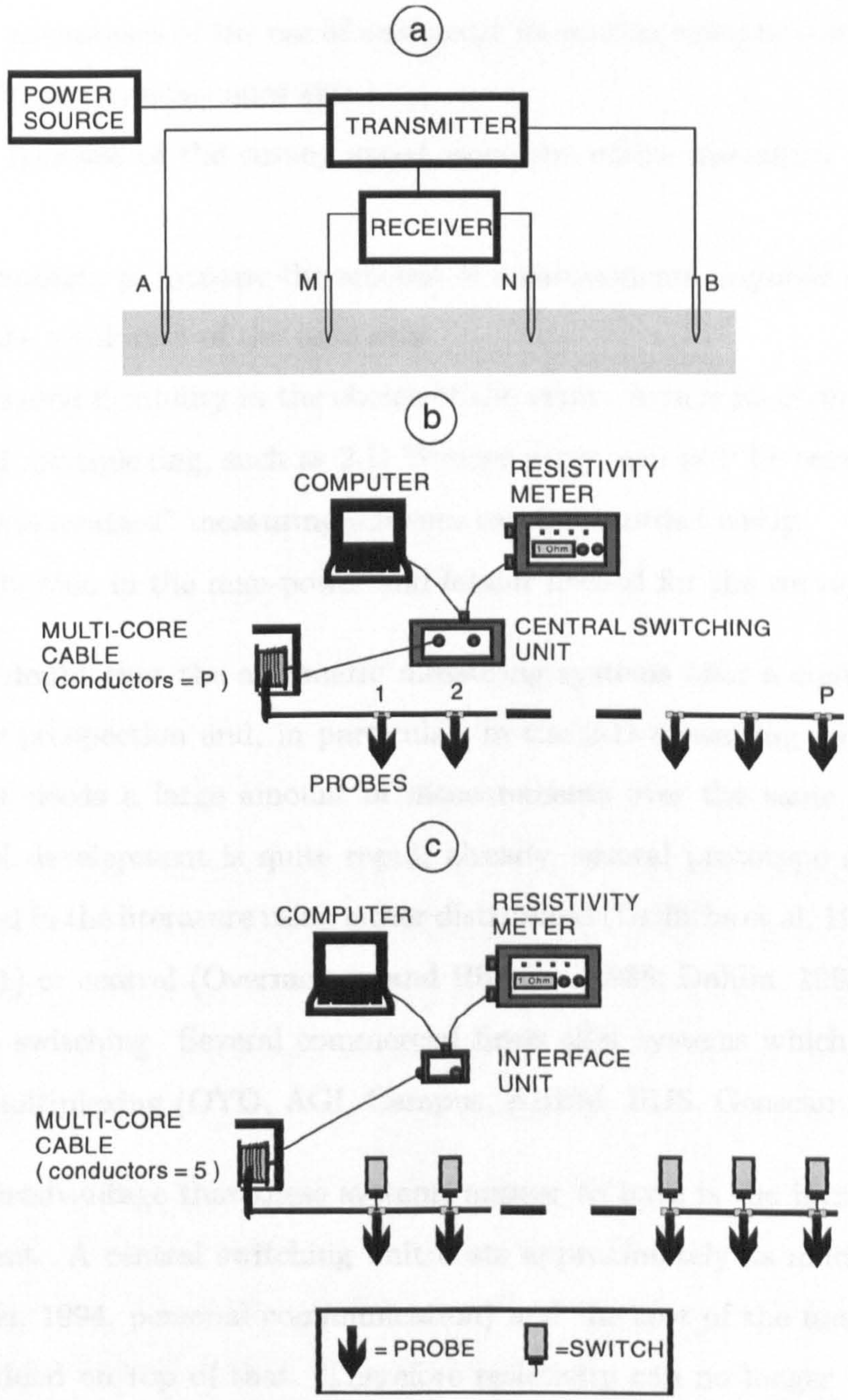


Figure 2.10: a) A block diagram of the main features of a resistivity meter. b) Principle system lay-out of a central switching computer-controlled resistivity measuring system. c) Principle system lay-out of a distributed switching computer-controlled resistivity measuring system.

(Dahlin, 1993).

The obvious advantages of the use of automatic measuring systems compared to the use of traditional (manual) ones are:

a) An increase of the survey speed since the entire measuring procedure is automated.

b) The ability to increase the amount of measurements acquired and therefore potentially the resolution of the data sets.

c) Increased flexibility in the choice of the array. Arrays involving quite laborious manual multiplexing, such as 2-D Wenner array, can now be easily measured. Further, "non-standard" measuring schemes can be recorded easily.

d) Reduction in the man-power and labour needed for the survey.

There is no doubt that the automatic measuring systems offer a significant boost in resistivity prospection and, in particular, in the 2-D measuring mode, which is the one that needs a large amount of measurements over the same region. The technological development is quite rapid; already, several prototype systems have been reported in the literature using either distributed (Griffiths et al, 1990; Noel and Walker, 1991) or central (Overmeeren and Ritsema, 1988; Dahlin, 1993; Kohlebeck et al. 1993) switching. Several commercial firms offer systems which can support automatic multiplexing (OYO, AGI, Campus, ABEM, IRIS, Geoscan research).

The main disadvantage that these systems appear to have is the increased cost of the equipment. A central switching unit costs approximately as much as the RM itself (Dahlin, 1994, personal communication) and the cost of the multi-core cable has to be added on top of that. Therefore resistivity can no longer be viewed as such a cheap technique.

2.2.5 Problems with resistivity surveys

Apart from the accuracy of the RM itself, there are several other sources of noise that could affect the quality of the field apparent resistivity data sets. Some of those are:

Electromagnetic coupling Whenever the transmitted current is changed, electromagnetic (EM) coupling between the transmitting and receiving cables will occur (Ward, 1990). The coupling increases with frequency, with cable length and with the earth conductivity. Arrays which have separate receiving and transmitting cables are less susceptible to electromagnetic coupling. Therefore, dipole-dipole or pole-dipole arrays are preferable to Wenner or Schlumberger arrays as far as EM coupling is concerned. For the automated measuring systems which use multi-core cables coupling can become a problem if the conductors in the cables are not screened.

Probe positioning Misplaced probes effect the accuracy of the measurements since they result in a miscalculated geometrical factor and consequently in a misleading apparent resistivity value. Extra care has to be taken in order to position the electrodes as accurately as possible. Automatic positioning instruments such as the electromagnetic positioning system can practically eliminate this source of errors with consequent increases in the survey costs. One other way to reduce positioning errors is to use probes mounted on a fixed frame. This approach is widely used in archaeological and pedological surveys where the small scale renders the use of a frame practical [8]

[8]One of these systems developed for the square array, widely known as RATEAU (Hesse et al., 1986), consists of four steel wheels, which act as a square frame. The frame can be towed by a vehicle which has an electromagnetic positioning facility - the system is capable of providing one reading per 10 cm and can cruise at a speed of about 15 miles/hour. Another system which uses probes mounted on a frame is the RM4 instrument (Geoscan Research). It is used mainly for archaeological surveys and can incorporate the twin-probe array. A modified version called RM15 can cope with a variety of arrays (e.g. Wenner, dipole-dipole etc).

Topography effects Topographic effects are important in resistivity surveying. In general topography causes dispersion and focusing of the current distribution and, in this way, artificial conductive and resistive anomalies. Slopes angles of more than 10 degrees create significant errors in measurements. Although only the behaviour of the dipole-dipole array has been studied, a way for correcting the topography effects (both for 2-D and 3-D effects) has been suggested (Fox et al., 1980; Oppliger, 1984). A further discussion of the correction of topographical effects will be made in Chapter 3.

Other sources of noise could be due to the geology and field conditions (surface inhomogeneities near the probes, overburden, excessive moisture etc.), or due to anisotropical behaviour of the subsurface. When interpreting the resistivity the level of noise has to taken into account in order to avoid artefacts in the interpretation results.

2.3 Measuring and Interpreting Resistivity Data

In this section the practical application of the resistivity technique is demonstrated. Further, ways for presenting and interpreting the data will be described. The way to perform a resistivity survey depends on the measuring mode that has been selected for the survey, and thus each resistivity measuring mode will be described separately.

2.3.1 Vertical electrical sounding

The purpose of the vertical electrical resistivity sounding (VES) procedure is to investigate the variation of subsurface resistivity with depth. This is achieved by arranging the measurement lay-out in such a way that for every consecutive measurement the measured potential difference is affected by the formations that lie at increasingly greater depths: the spacing of the current probes is increased systemat-

ically. The entire procedure is based on the assumption that the subsurface consists of distinct horizontal layers which are homogenous and isotropic.

Application

In this procedure, only electrode arrays with internal symmetry (no probes at infinity) are used: Schlumberger, Wenner, dipole-dipole and square array since at every measurement the separation of the electrodes is increased relative to a fixed centre O (which is the centre of the MN for Wenner and Schlumberger arrays, the mid-point of the two dipoles for the dipole-dipole array and the centre of the square for the square array). In order to achieve good resolution with depth the array should be expanded about six times per decade of distance (in a logarithmic scale).

The Schlumberger array is used most commonly (Rijo et al., 1974): due to the assumption that the distance between the current probes is much larger than that of the potential probes ($AB/2 > 5MN$), the potential difference between MN represents the electric field at the mid-point between A and B. At every measurement the current electrodes are displaced outwards (in order to change the depth range) while the potential probes are left at the same position. However, when the ratio between AB and MN distances becomes very large (20-50) the MN distance has to be increased as well, otherwise the measured signal is two weak.

The Wenner array is also used for VES. The application is similar to that for the Schlumberger array, but for every measurement the potential probes have to be moved as well - this is not a problem, however, when automatic multiplexing is used. The Wenner array has the advantage that due to the large MN separation the measured signal is quite high (see Table 2.3); but for the same reason, however, Wenner measurements have an unpredictable level of noise caused by the existence of lateral inhomogeneities [9] (Koeford, 1979).

[9]The effect of the lateral inhomogeneities on the Wenner sounding configuration was studied by Barker (1981) who proposed an efficient way for taking them into account while measuring (offset

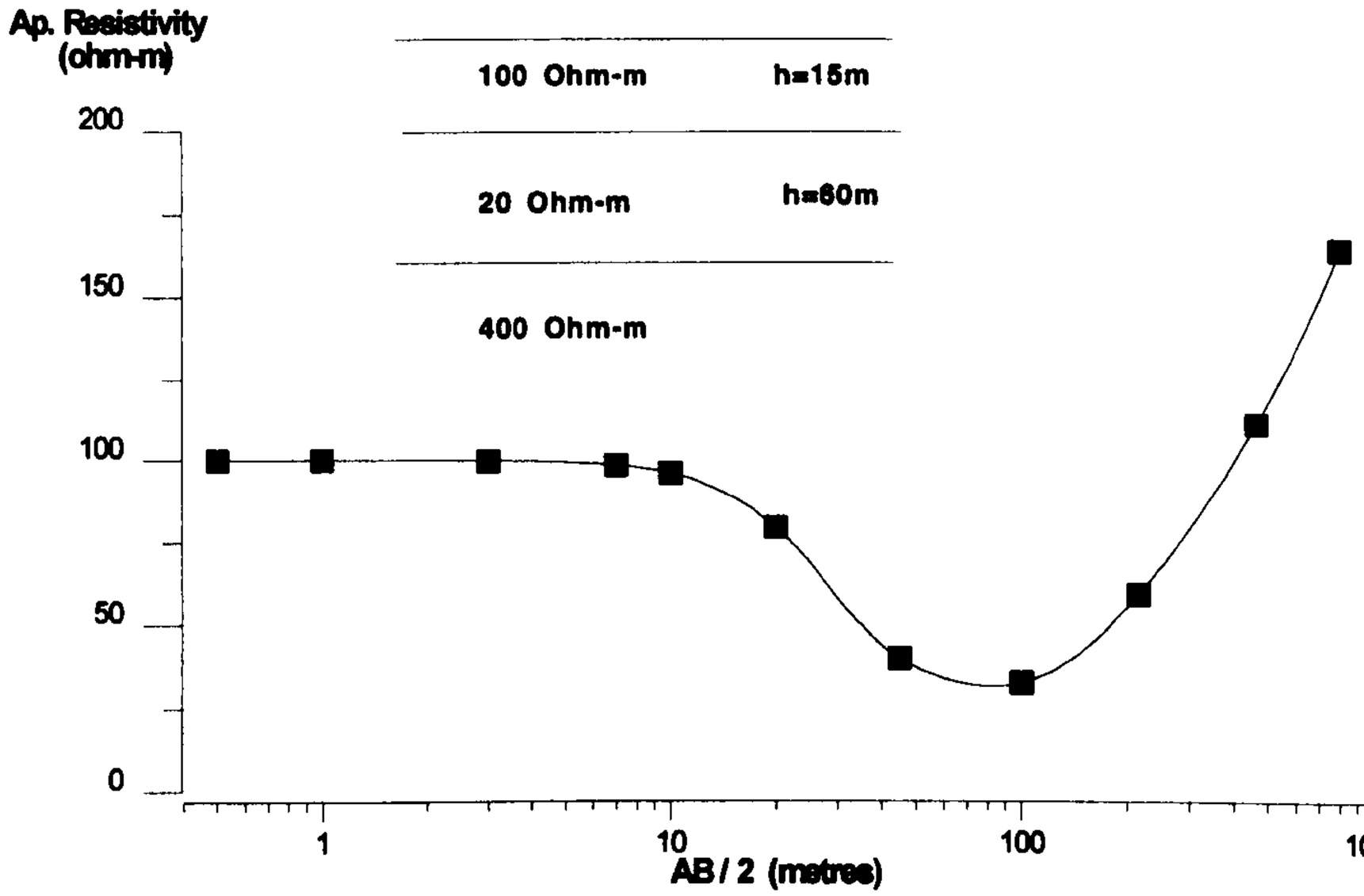


Figure 2.11: Schlumberger VES data set for a three layered earth model(top).

Dipole-dipole VES can be used for very large scale exploration since it has favourable cable logistics (Al'pin, 1950) but on the other hand due to its low signal is not widely used . Finally, square array VES is mainly used for ascertaining directional variations in the subsurface resistivity (Al Hagrey, 1994)

Vertical soundings are applied extensively for stratified earth studies, since stratification is a characteristic that is usually relatively constant over considerably large areas. Therefore a sparse grid of soundings can be enough to deduce useful conclusions about the subsurface. It follows that with soundings it is very difficult to get an indication of the lateral variations of resistivity. Due to the importance of the hydrogeological applications of soundings there is a large amount of research being conducted in this field (Koeford, 1979).

Wenner method).

Interpretation

The VES curves are presented in logarithmic scale. A typical Schlumberger sounding curve for a three layer earth is shown in Figure 2.11. In VES the basic assumption is that of an earth with resistivity varying only in one dimension (depth), and therefore the resulting interpretation is called 1-D. The interpretation of the VES curves is traditionally made by matching parts of the measured curve (the continuous curve produced by interpolation of the discreet measured VES data points) with master curves produced for two or three layered earth models (Koeford, 1991). Although nowadays automatic full-curve matching (inversion) schemes are used regularly (Inman, 1973; Zhody, 1989) manual curve matching is still useful for giving the interpreter a "feel" for the data, or for providing initial models to be used in automatic interpretation schemes.

There are several factors which can affect severely the validity of the interpretation of the VES data: The existence of severe lateral inhomogeneities (departure from 1-D model), non horizontal layer interfaces, anisotropy etc. Moreover the assumption of an earth which consists of distinct homogenous layers is not always valid since changes of the resistivity with depth can be so gradual and continuous that no distinct layer units can be assumed (Constable et al., 1987). Finally, an extra problem is imposed by the principle of equivalence (Kunetz, 1966): for some specific layer sequences there are several layer models which produce almost identical VES data sets. All of these factors have to be taken into account when measuring and interpreting VES data.

2.3.2 Lateral profiling

In the lateral profiling (LP) procedure the separation of the probes remains stable throughout the survey while a series of measurements are taken by moving the array in a lateral direction. By plotting the resistivity values along the measured profile

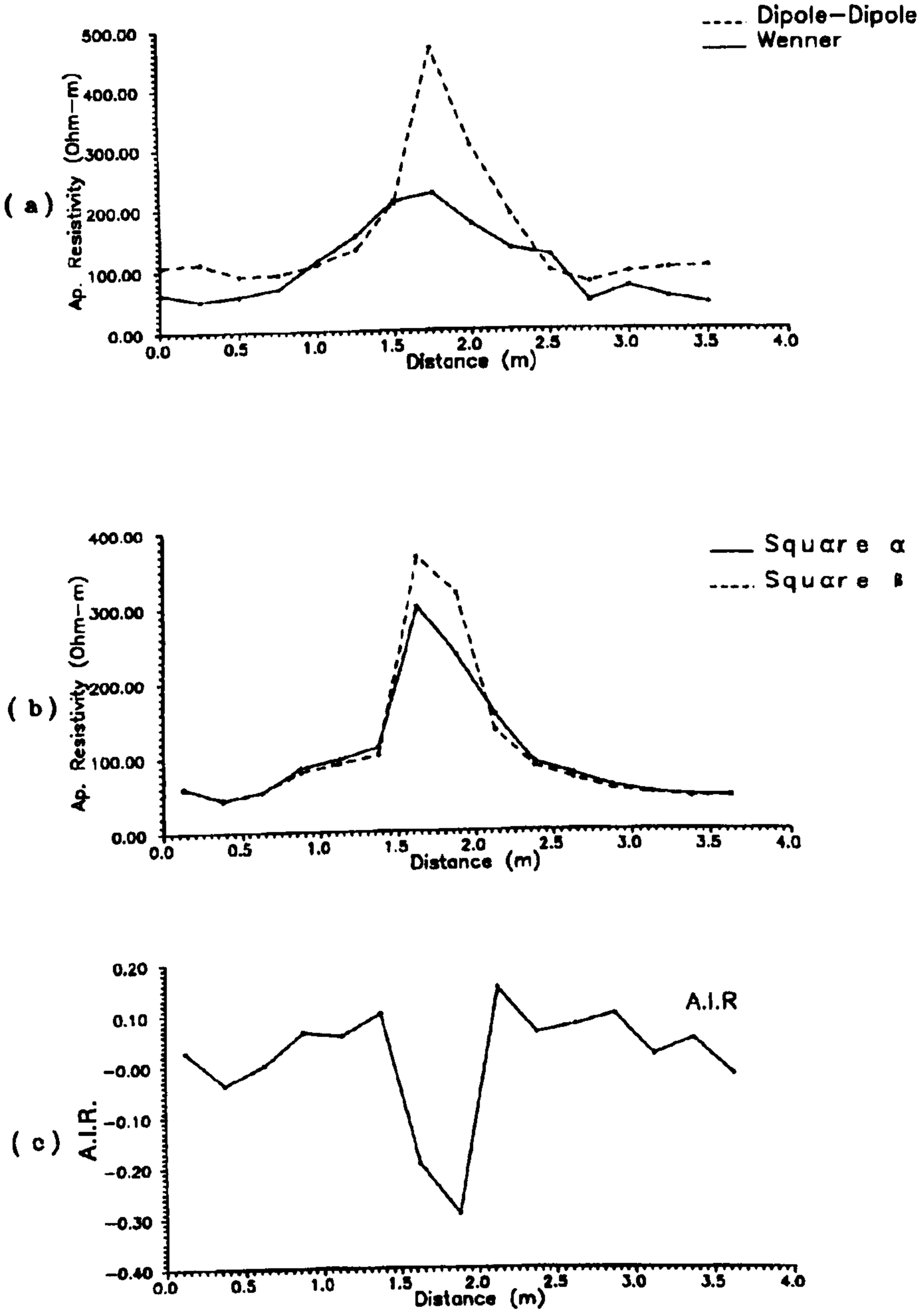


Figure 2.12: Apparent resistivity profiles with different arrays over a buried wall

the anomaly caused by a subsurface feature can be picked out.

In profiling, the lateral resistivity anomalies such as dykes, faults and walls can be picked up. However, the information is restricted due to the stable penetration depth which is imposed by the constant separation of the current probes. In practice, most of the time the measurements acquired "represent" the upper part of the feature, therefore any attempts to describe the total shape of the feature are based more on prior knowledge.

Application

There are two ways for employing lateral resistivity profiling, depending on the size of features one is interested in. Profiling of geological features takes place in separate sections with large probe separations and is used mainly for reconnaissance purposes.

Conversely, when it comes to more detailed surveys (archaeological applications), a different approach in profiling is followed. A dense grid of parallel sections is created and the resistivity measurements acquired at this pattern are plotted to the exact grid coordinates. In this way a resistivity map is created which can lead to the identification of various features such as walls, ditches, tombs etc.

Any of the various arrays can be used for LP. From a practical point of view the scale of the survey is essential: for quite deep surveys where a large probe separation is needed, the arrays with remote probes (probes at "infinite" distance) are not practical. Consider, for instance, the twin-probe array with a separation of 50 meters. The remote probes should be placed (50x30) 1500m away from the active ones- a distance which is not at all practical. On the other hand, for shallow civil engineering or archaeological targets, these arrays are far more preferable since less probes are being moved with each measurement.

The choice of the array depends also on the sensitivity of the array to the existing

lateral inhomogeneities (See Table 2.3). The array that is most widely used in the LP procedure is the gradient array (which was designed for profiling purposes only). Pole-pole is also quite popular while dipole-dipole and pole-dipole are used as well. Wenner and Schlumberger arrays are not widely used for LP.

Interpretation

The interpretation of the LP data is related to the shape of the various anomalies in the data set. In general the interpretation is made by assuming that the array will produce high apparent resistivity anomalies over resistive features and low ones over conductive bodies. [10].

Arrays that give anomalies which have a width approximately equal to the width of a feature and a clear maximum which corresponds to the feature resistivity distribution are considered to be ideal. In general, the shape of the resulting anomaly is a complex combination of effects, due to the geometry of the array and the characteristics of the body. In particular the anomaly is affected by: a) the position of the array relative to the feature, b) the resistivity contrast between the body and the surrounding environment, c) the depth that the body is buried.

As long as the proper array is chosen, the interpretation of LP results can be easy and quite accurate. This can be seen in the examples presented here: In Figure 2.12 apparent resistivity profiles with different arrays over a wall are shown. For every array the anomaly produced is quite pronounced. Moreover in Figure 2.13 (Szymanski et al., 1992) and in Figure 2.14 (Tsokas et al., 1994) examples of recent LP surveys using a twin-probe array are shown. The distribution of the apparent resistivity (shown here in a grey scale) clearly indicates the location of the buried features.

[10]This is broadly true but (due to the variable sensitivities that the arrays have) for specific locations of the array relative to the body the opposite effect can occur (negative sensitivity). A more detailed discussion on that feature is presented in chapter 4

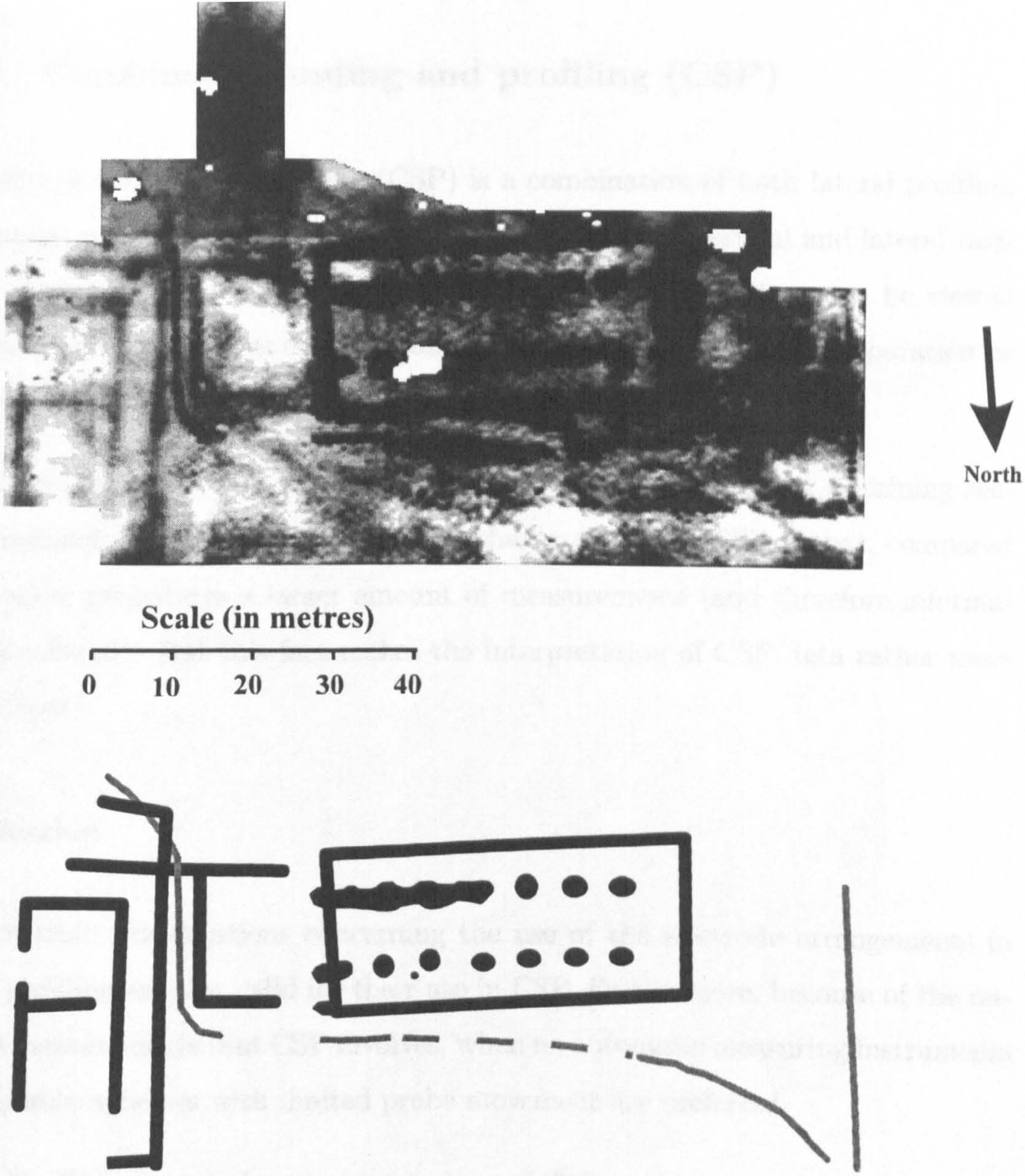


Figure 2.13: Grey scale resistivity map (twin-probe, 0.5m spacing) from the site of Fountains Abbey (arbitrary gray scale - contrast equalise. d) The pattern of the Abbey's guest house can be very easily seen. (Szymanski et al., 1992).

2.3.3 Combined sounding and profiling (CSP)

Combined Sounding and Profiling (CSP) is a combination of both lateral profiling and vertical sounding procedures: information about both vertical and lateral variations of resistivity can be acquired. The data acquisition pattern can be viewed either as a series of profiles over the same area with increasing probe separation or as a series of adjacent vertical soundings along the profile line.

Almost all arrays that can be used in profiling can also be used for obtaining sectional resistivity images. One of the main characteristics of CSP is that, compared to the other procedures a larger amount of measurements (and therefore information) is collected. But this fact makes the interpretation of CSP data rather more complicated.

Application

The practical considerations concerning the use of the electrode arrangements in lateral profiling are also valid for their use in CSP. Furthermore, because of the numerous measurements that CSP involves, when no automatic measuring instruments are available, schemes with limited probe movement are preferred.

Hence, the Wenner array is not extensively used if there is not an automatic measuring procedure, while for geological applications dipole-dipole is preferable (Stretenovic et al., 1992). When it comes to shallower applications apart from dipole-dipole (Patella, 1978), either the pole-dipole or pole-pole array can be employed.

When a dipole-dipole array is used, the length, a, of every dipole, (AB, MN), remains stable and the same profile is repeated by increasing the separation, na, between the dipoles by an integer number of dipole lengths (n=1,2,3...). In practice, in order to minimize the probe movements one dipole remains stable (usually the measuring one) and the other is moved along the profile. As soon as the measuring sequence is completed a new one starts, but this time the stable dipole is moved one position

Figure 2.14: Grey scale resistivity map (twin-probe) from the Classical/Roman site of Europos (N.Greece). The high anomalies (dark) represent the ruins of the ancient acropolis (Tsokas et al, 1994).

forward (see Figure 2.15a). The maximum dipole separation n_{max} is usually no larger than 6-7 dipole lengths since for larger values of n the measured signal is usually too low and the quality of the measurements drops. The number of measurements (N_{dd}) acquired by this procedure for the dipole-dipole array, if P is the number of probes will be:

$$N_{dd} = \sum_{i=1}^{n_{max}} (P-2) - i \qquad n_{max} < P-2 \tag{2.46}$$

A similar measuring scheme can be followed when the pole-dipole array is used for CSP. This time instead of the current dipole a single current probe exists, and that fact increases the number of the acquired measurements (N_{pd}).

$$N_{pd} = \sum_{i=1}^{n_{max}} (P-1) - i \qquad n_{max} < P-1 \tag{2.47}$$

A similar approach can be followed for the pole-pole CSP. It is interesting to note that if an ideal pole-pole array [11] CSP data set is obtained, then all of the measurements that would have been produced by other arrays (dipole-dipole, pole-dipole, Wenner) can be produced (using superposition) from that data set. The number of measurements (N_{pp}) produced with the pole-pole CSP is given by

$$N_{pp} = \sum_{i=1}^{n_{max}} P - i \qquad n_{max} < P \tag{2.48}$$

The Wenner array CSP is obtained somewhat differently. The entire profile is obtained for a stable probe separation n=1a. This procedure is repeated for every probe separation until $n = n_{max}$ (see Figure 2.16a). The total number of measurements (N_{wn}) obtained is given by

$$N_{wn} = \sum_{i=1}^{n_{max}} P - 3i \qquad n_{max} < P/3 \tag{2.49}$$

The ways for obtaining CSP data sets presented here apply only when manual multiplexing is used. When automatic measuring systems are available there is more flexibility in the way the measurements can be obtained [12].

---

[11] B,N probes at infinity

[12] Xu (1993) proposed a measuring scheme which involves all the possible independent measurements for a number of surface electrodes. The scheme turns out to be a combination of a full dipole-dipole data set with several Wenner-like measurements.

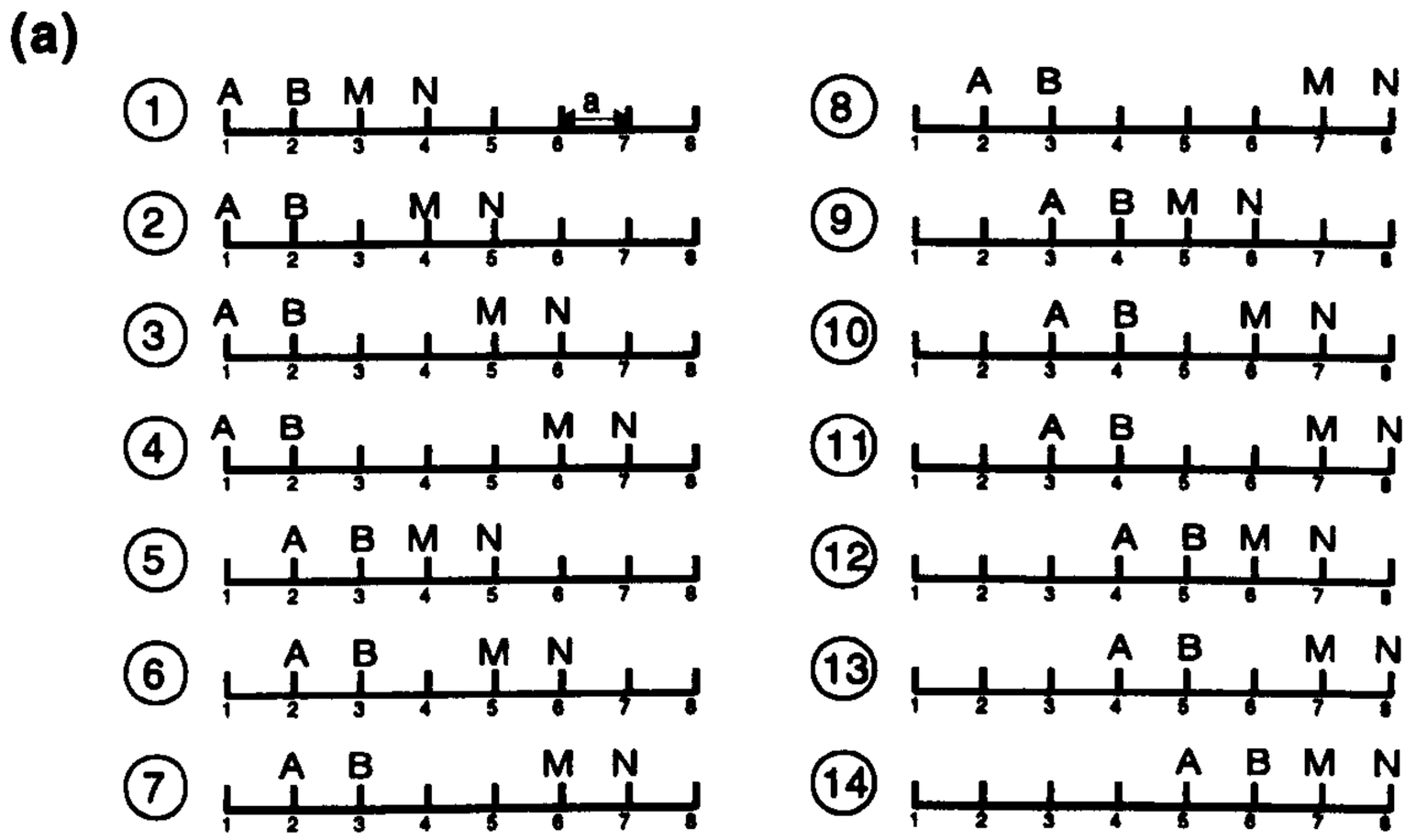
One of the problems of the CSP procedure is that the common use of equally spaced probes does not allow adequate sparsity of measurements and therefore high resolution of horizontally layered structures. If high resolution with depth is needed the survey could be carried out using non-integral values of n but this is not practical. In such a case a better choice is the use of Schlumberger CSP: A survey line with equally spaced points is chosen and a full VES data set is obtained having as a centre each of these points. Each sounding is expanded in a direction parallel to the survey line.

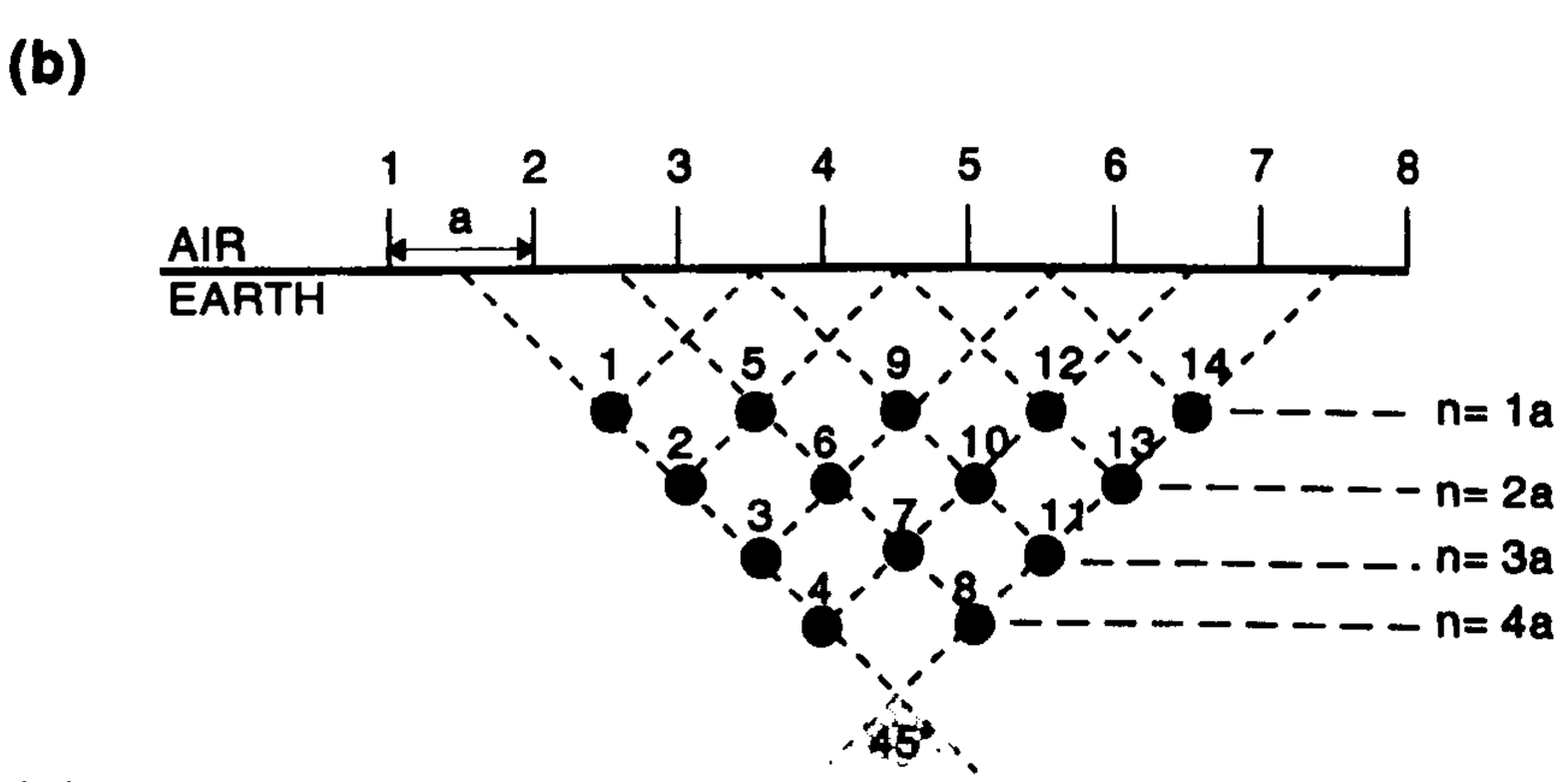
Theoretically, CSP is the procedure that guarantees the maximum amount of information in resistivity prospecting. This advantage is counterbalanced by the fact that it is the most laborious when compared to VS and LP in isolation. Therefore traditionally it is used either when detailed information about the structures sought is needed or when the other methods fail to give reliable results. In both cases it is supplementary to the basic VS and LP procedures. However, the recent development of computer driven measuring systems has given a boost to its use, and so it is believed that, at least for shallow depth applications, it will in future be used as a stand-alone measuring procedure.

Interpretation

The CSP data sets are presented traditionally using the pseudosection method. It was proposed initially by Hallof (1957) for the dipole-dipole data set. It is based on the fact that the bigger the dipole separation, the more the resistivity is related to greater depths. Hence, each measured resistivity value is arbitrarily placed at the intersection of two 45-degree lines through the centres of the dipoles (see Figure 2.15b). In this way an approximate 2-D sectional resistivity image (pseudosection) which is easy interpretable is produced.

For the Wenner array the pseudosection is constructed by projecting each apparent resistivity measurement to the midpoint of the MN probes and at a depth equal to





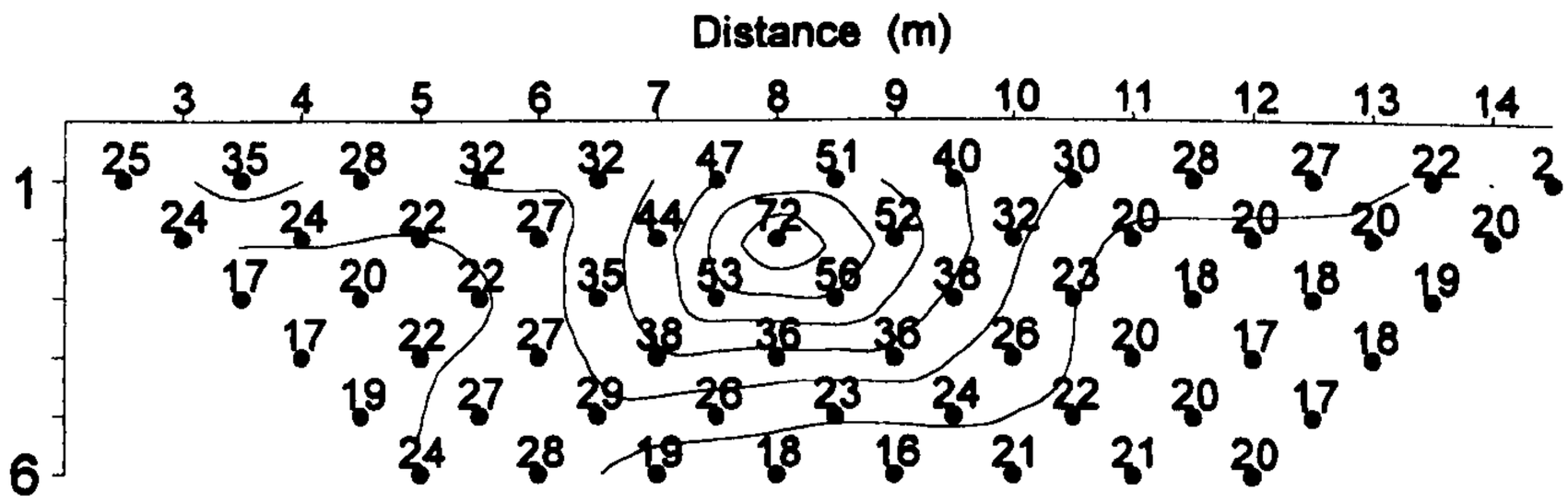


Figure 2.15: Measurement scheme for the dipole-dipole CSP measuring mode: a) The way the data set is acquired for 8 electrodes and maximum separation n=4. b) The representation of the data set in a pseudosection form. Each number corresponds to the number assigned to each measurement in Figure a. c) Dipole-dipole pseudosection of a data set acquired at the archaeological site of Europos (N. Greece).

the inter-probe spacing of that measurement (see Figure 2.16b). In a similar fashion pseudosections for the pole-dipole and pole-pole arrays can be constructed.

Edwards (1977) suggested that changing the depth scale of the pseudosection results in more geologically realistic images of the earth. The scale that he suggested for each array can be seen in Table 2.5. Pseudosection images can be presented either in the form of contour maps or with the form of colour or grey scale images. A detailed presentation of the pseudosection technique is given is chapter 5.

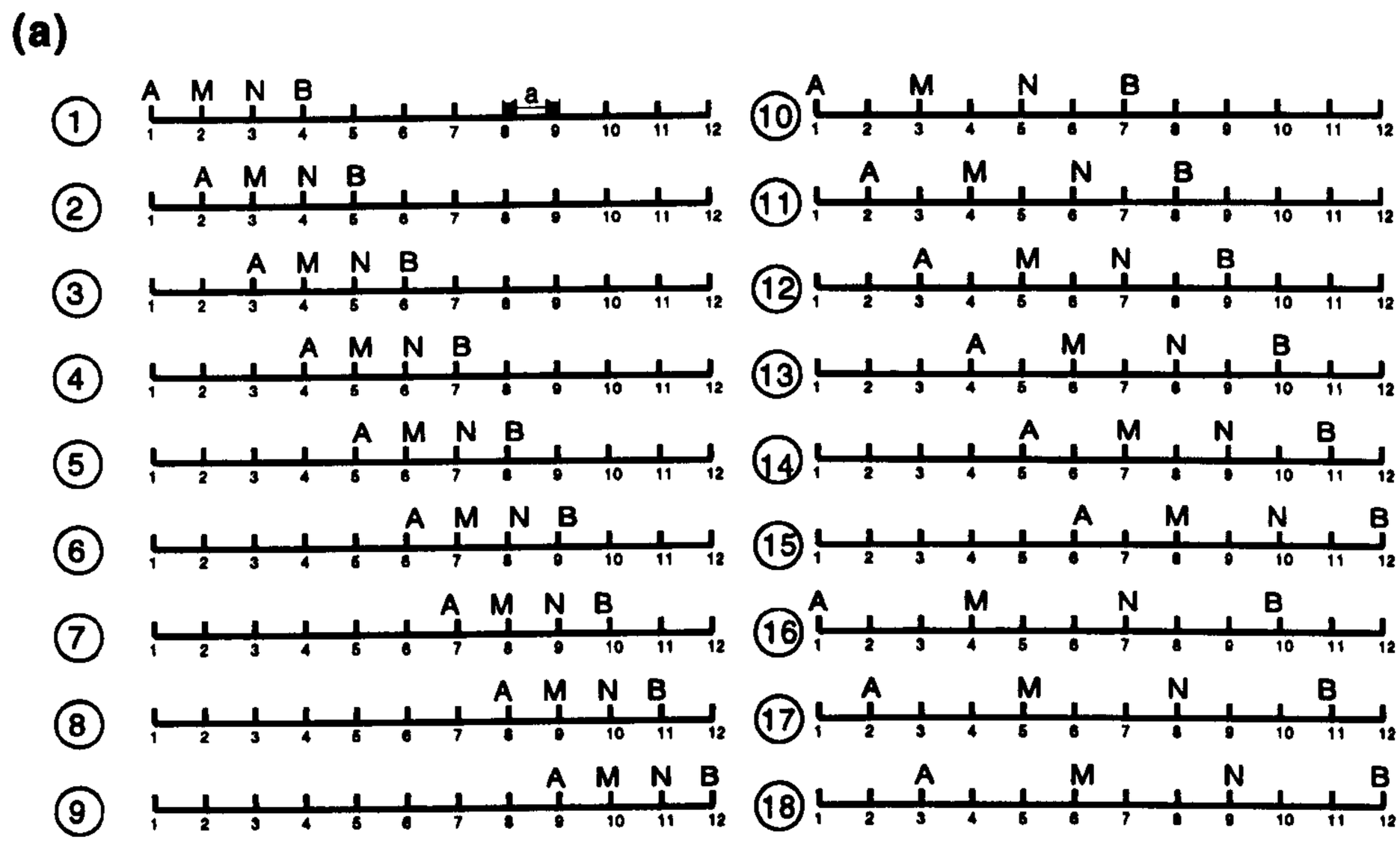
More recently the interpretation of the CSP data is carried out made by using semi or fully automatic modelling and inversion schemes which will be presented analytically in chapters 5 and 6.

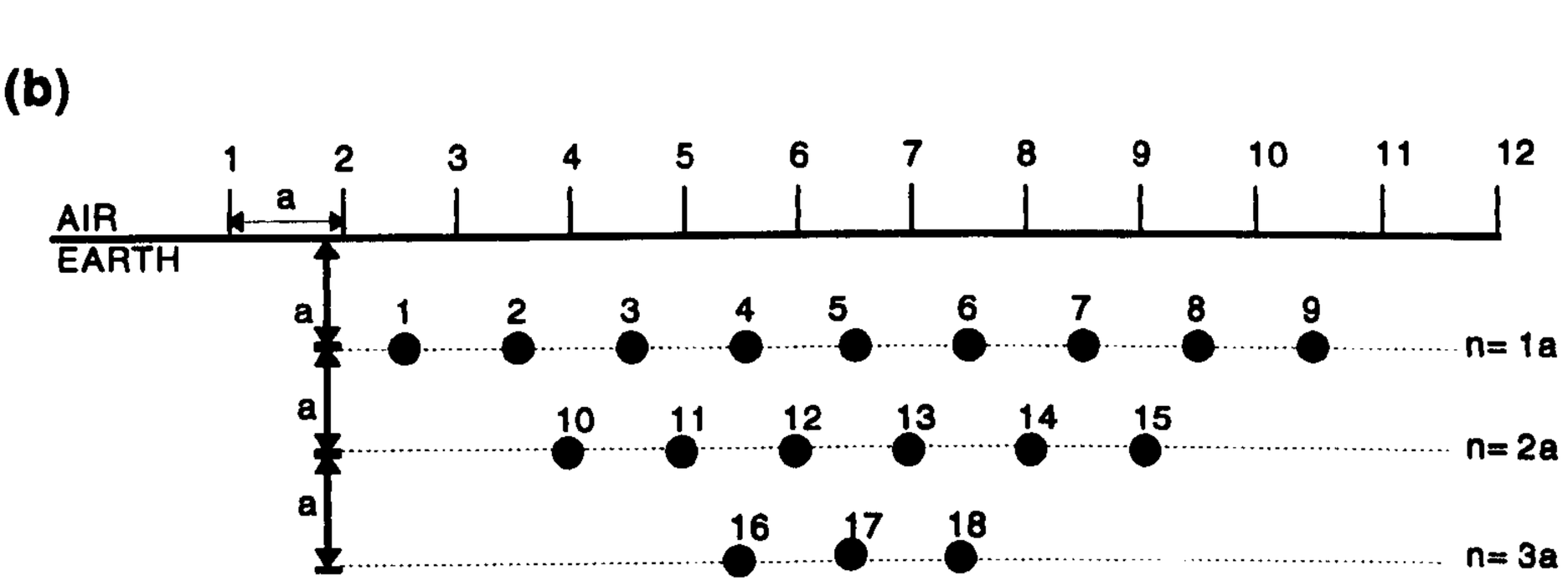
In Figure 2.15c a pseudosection (in a contoured map form) of a dipole-dipole data set obtained from the Archaeological site of Europos (N.Greece) is presented. The survey line was positioned over a buried tomb (previously identified by a twin-probe LP survey) which gave the high apparent resistivity values positioned at the centre of the pseudosection.

In Figure 2.16c a pseudosection (in a contoured map form) of a Wenner data set obtained at the courtyard of the Electronics Department at York is presented. The survey line was positioned over a drainage cavity. The high resistive structure can be clearly identified in the figure.

Naming

There are several other names that are used to describe the CSP procedure. The most common name is pseudosection (from the Greek "pseudo", meaning in this case approximate). This name describes mainly the interpretation method (which gives an approximate section of the subsurface) and not the measurement procedure itself.





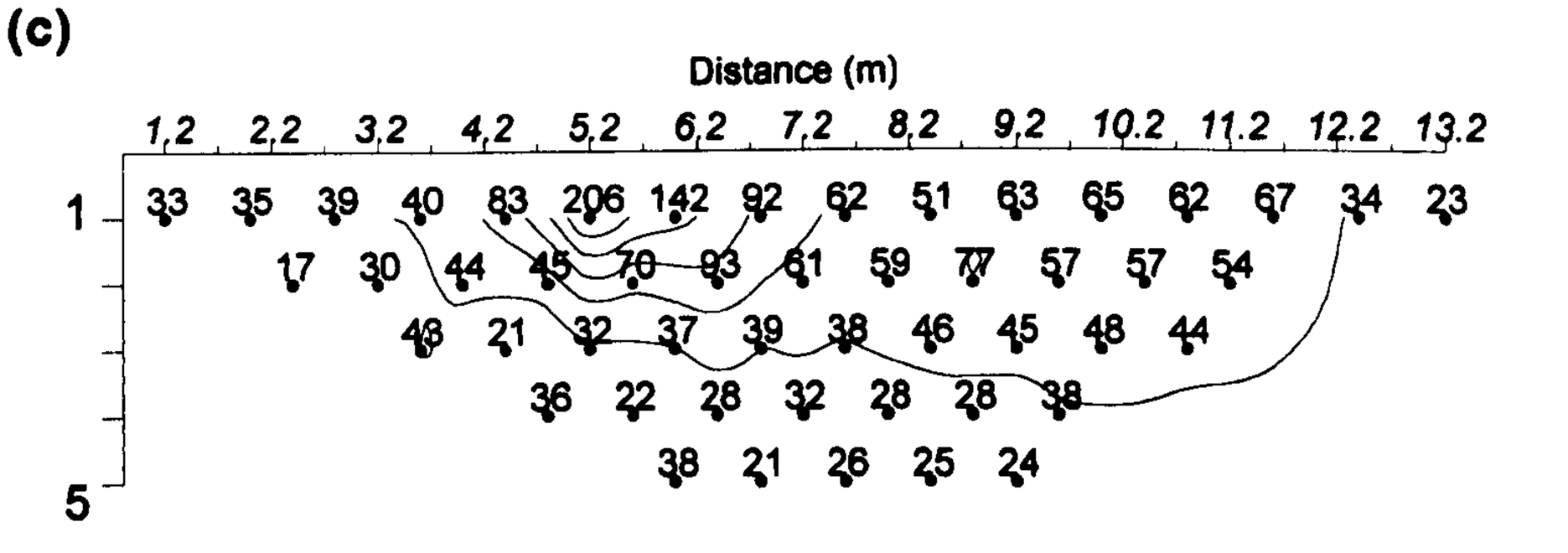


Figure 2.16: Measurement scheme for the 2-D Wenner array: a) The way the data set is acquired for 12 electrodes and maximum separation n=4. b) The representation of the data set in a pseudosection form. Each number corresponds to the number assigned to each measurement in Figure a. c) Wenner pseudosection of a data set acquired at a courtyard of the Electronics department at York over a drain.

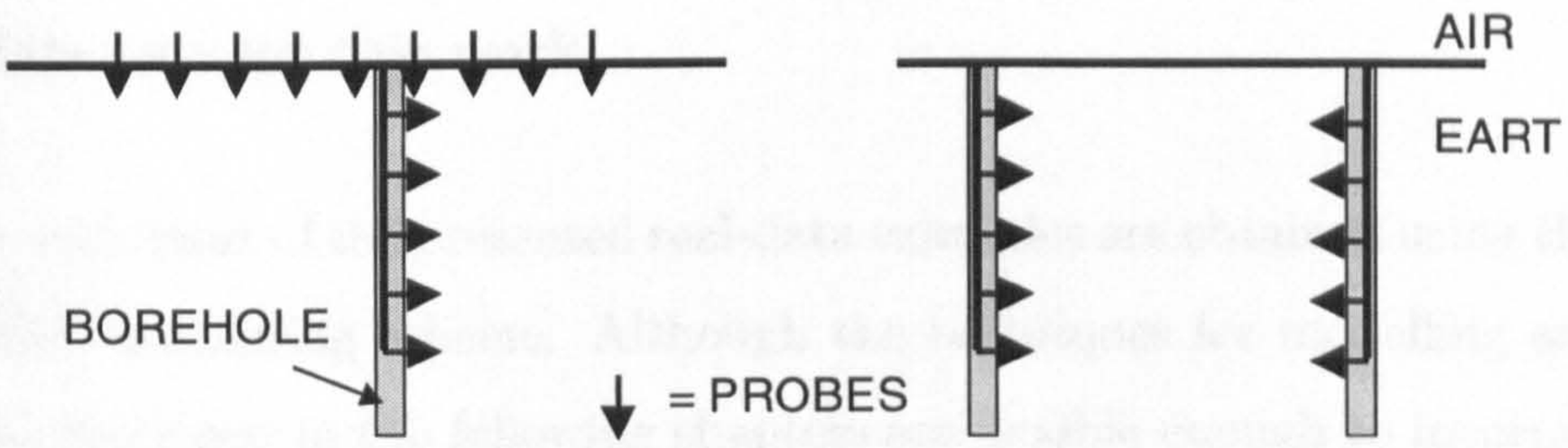


Figure 2.17: Borehole-to-surface and borehole-to-borehole probe arrangements.

In general it is quite common to associate the measuring mode with the interpretation: several times the VES procedure is described as a 1-D resistivity survey and in the same context the CSP procedure is described as a 2-D resistivity survey. In this work, however, we will use the 2-D term strictly for the interpretation (inversion) schemes.

The name resistivity tomography is also used: the term is due to the similarities of the procedure to medical imaging techniques (X-ray tomography). This term was first used by Shima in order to describe the borehole-to-surface and borehole-to-borehole resistivity measuring modes (see Figure 2.17) which although different in physical terms, can be treated similarly as far as the measuring configuration is concerned: a dipole-dipole CSP measuring sequence for surface-to-surface data can be easily transferred to accommodate borehole-to-surface measurements.

On the other hand, the automated measuring systems allow absolute flexibility in the choice of the probes and the arrays (Xu and Noel, 1993), so in a sense the term tomographic is more general than the term CSP. In that context the term "tomographic resistivity data sets" will appear in this work and also embraces traditional CSP data sets. This is because the modelling and most of the reconstruction techniques that will be described can be applied to *any* sort of electrical data (surface, borehole-to-surface, borehole-to-borehole).

The data sets for this work

For this work most of the presented real-data examples are obtained using the dipole-dipole CSP measuring scheme. Although the techniques for modelling and reconstruction described in the following chapters are flexible enough to incorporate any short of data set the dipole-dipole CSP was chosen for two main reasons:

a) Since an automatic measuring system was not always available an easily applicable array should be chosen (minimum probe motions).

b) The comparison of the resistivity arrays showed that this array has good resolution (both lateral and vertical). Further, recent studies (Sasaki, 1992) showed that dipole-dipole can produce inversions with a higher resolution compared to the other arrays.

c) The vast majority of research work on the 2-D reconstruction of resistivity data has been done using this array. Hence, direct comparisons can be made.

2.4 Chapter Overview

In this chapter the basic resistivity theory has been reviewed. The basic mechanism and equations that govern the flow of the electrical current into the earth were presented. Basic concepts such as that of the apparent resistivity, the reciprocity and the anisotropy were explained. Further, the artificial nature of the apparent resistivity was demonstrated and a review of the existing resistivity arrays showed some of their merits and limitations.

The resistivity instrumentation and the automated systems were presented and the practical application of the resistivity approach was explained. For the combined sounding profiling mode which is one of the interest of this thesis it was decided that the dipole-dipole array was the best choice in view of the its merits and the hardware limitations.

Chapter 3

Forward Resistivity Modelling Using the Finite Element Method

In this chapter the forward resistivity problem is addressed. The existing approaches for modelling earth-resistivity data are discussed and the reasoning for using the finite element method (FEM) is presented.

Although a detailed analysis of the application of the 2.5-D FEM into the resistivity modelling is given in many works, the scheme developed for this work is described in full. This is necessary because full justification of techniques that are presented in this work is possible only if the core of the 2.5-D FEM modelling is explained fully. The computational aspects and accuracy of the proposed scheme are addressed.

Further, a strategy for modelling arrays parallel to the strike direction (i.e. square array) is proposed. Finally the scheme is used to study the effects of terrain topography on commonly used resistivity arrays.

3.1 Resistivity Modelling

Forward resistivity modelling, (FRM), is the procedure of obtaining the surface potential response of a model of a known resistivity distribution, interacting with a particular applied current. In other words, the FRM involves the solution of the equations that govern the flow of electrical current through the inhomogeneous ground for a specific resistivity distribution and current source configuration such that the potential distribution can be found. The potential distribution could be calculated either at only the measuring points, or at a range of points over a specified area. Once the potential is known it is straightforward to calculate the apparent resistivity. There are two different ways of handling the solution of the forward problem:

Analytical approach: this involves the direct solution of the field equations. In practice, only very simple structures,such as a buried sphere (Cook and Van Nostrand, 1954), have known solutions. Analytical solutions, although of not any direct practical significance, are useful for validating modelling results produced by other techniques.

Numerical approach: this involves mathematical procedures based on arithmetic operations performed by computers. This implies that a way must be found for expressing the solution of the continuous field equations at a finite number of discrete points. The numerical methods have the great advantage of being able to incorporate arbitrary resistivity distributions, and therefore they are preferable to analytical ones.

There are a variety of numerical modelling techniques that have been applied for solving the forward resistivity problem. These can be separated into two main categories:

Integral equation methods: the area where the solution is sought is restricted to the surface of resistivity discontinuities. The principle of the numerical

calculation is based on the use of a Green's function, which in a broad sense assumes that the existence of a perturbing body of different resistivity from the background is equivalent to a distribution of elementary electrostatic sources on its surface. In order to decide the distribution of the potential at the earth's surface the cumulative effect of all of the elementary sources is calculated by integration and is added to the potential distribution caused by the source. The integral method was mainly developed for handling simple resistivity distributions (Keller and Frischknecht, 1966; Lee, 1974; Furness, 1992). More recent works incorporate inhomogeneity within layered media and give numerical checks (Das and Parasnis, 1987).

Differential methods: the area of the solution is discretized and is constrained by the introduction of artificial boundaries to the problem in order to be compatible with computer finite arithmetic. The solution is calculated at every discrete point of the study area. Hence, a different resistivity can be attributed within each discrete fragment, so that differential methods can cope with any resistivity distribution, no matter what its complexity.

By comparing the two methods it can be said that integral equation methods have the advantages of being less time-consuming and occupying less computer memory than differential methods since the solution is confined to the surfaces of the discontinuity rather than through the entire area of study. Conversely, differential methods have the attractive advantage of handling complicated resistivity distributions. Moreover their theoretical development is less laborious than that for the integral equation methods.

The choice of which method to use is closely related to the purpose of the forward model: integral equation methods are preferable either when the intention is to study the potential distribution at the surface caused by simple structures, or to study the responses caused by moving different arrays over the same body (Bernabini et al., 1987). On the other hand, when the final goal is to incorporate the

forward model within an iterative inversion scheme in order to decide a resistivity distribution based on real measurements, the choice should be clearly a differential method (Hohmann, 1988) since the subsurface resistivity distribution can be complicated. Moreover, during the inversion procedure not only the apparent resistivity (AR) measurements but also the first derivatives of the AR in respect of changes in the resistivity of discreet blocks (distributed all over the studied space) have to be calculated. Differential methods can easily cope with these calculations since they provide a solution for the entire studied area. Hence, no further justification is needed for the choice of a differential method for this work.

3.1.1 Modelling using differential methods

Three differential forward resistivity modelling schemes are reported in the literature. These are: resistor network analysis, finite difference, and finite element methods. They all involve the calculation of the potential distribution (due to point sources) at discrete points of a medium of arbitrary resistivity. The fact that the domain is restricted by imposing artificial boundaries to an open boundary problem (semi-infinite space) signifies that a unique solution can be obtained only if knowledge concerning the behaviour of the potential and its derivatives at these boundaries is available (boundary conditions). A short presentation of these methods will follow in order to pinpoint their similarities and differences:

Resistor Network Analysis Method (RNAM) - This method is based on the direct analogy between field equations and the equations governing voltage about a point source in resistance networks. The subsurface is discretized into a resistor network; thus the resistors are being used to simulate the earth's resistivity. The potential is decided at the nodes by solving a system of linear equations (which is constructed via Kirchhoff's second law) :

$$\mathbf{L}\ \mathbf{V} = \mathbf{S} \tag{3.1}$$

where:

L is a sparse banded matrix which depends on the features of the network.

V is the unknown potential vector.

S is a vector describing the current sources.

The method was incorporated in an inversion algorithm suggested by Pelton et al. (1978) and Tripp et al. (1984) . The biggest advantage of the method is that it allows a physical insight into the simulation of the earth, unlike the other two methods which follow a more theoretical approach.

Finite Difference Method (FDM) - The continuous medium is divided into rectangular cells, each associated with a point to which a resistivity value is attributed. Hence, a grid of discrete points is formed at which the unknown potential can be computed. The partial derivatives of the governing equation (see section 2.1.3) are approximated by evaluating them at a point as a central difference of the potential at its neighbouring points. Further, taking into account the boundary conditions for the exterior points, the solution for the unknown potential can be obtained by solving a system of linear finite difference equations which has the form :

$$\mathbf{C}\ \mathbf{V} = \mathbf{I} \tag{3.2}$$

where:

C is the sparse banded matrix (called the coupling matrix) dependant on the resistivity distribution and the coordinates of the points.

V is the unknown potential vector.

I is the vector describing the current sources.

The FDM was first developed for resistivity modelling by Mufti (1976) and Dey and Morisson, (1979a,b). Ever since, it has become a very popular modelling technique, mainly because of the relative simplicity of its theoretical formulation.

Finite Element Method (FEM) - In the FEM the field is subdivided into elements, i.e. into subregions where the unknown potential is approximated by suitable

interpolation functions attributed to specific points of the element, which are called nodes. These functions contain as unknowns the values of the potential at the respective nodes of each element. The minimization of the residual caused by this approximation results in the formation of an algebraic system of linear equations (similar to the one produced by RNAM and FDM), the solution of which leads to the determination of the potential values at the nodes.

3.1.2 The choice of FEM for this work

The selection of a numerical modelling technique is based on several considerations, the most important of which are: computing time, computing memory, solution accuracy and flexibility.

By comparing the differential modelling techniques used for resistivity modelling it is quite clear that they all involve the solution of a system of linear equations having the general form of $K.V = F$, which will give the potential distribution V in the study area. According to Pelton et al. (1987) there is practically no difference in the size and type of the matrix K in all these methods. Since the solution of this algebraic system of equations is the most time-consuming procedure (more than 80%) it follows that (more or less) the time and memory needed is the same for all three methods.

Dey and Morisson (1979a) suggests that his FDM program appears to be faster than the FEM program (Coggon, 1971) when similar meshes are used; however this could be due to differences in programming efficiency. Furthermore, according to Fristiani et al. (1980) it is very difficult to set valid criteria for direct time comparison of the two methods since the design of the mesh used is based on different philosophies.

In terms of accuracy Rijo (1977) presented a simple example for which all three methods produced identical results when similar meshes were used. Although the example presented was for a limited case his conclusions can be easily generalized:

assuming similar mesh design strategies, it can be concluded that no method appears to show superiority as far as the accuracy of the results produced is concerned.

Thus, the choice of which method to use should be based upon the individual features that each method has. All methods can incorporate variable size meshes but only the FEM can cope with irregular shapes and boundaries [1], while the other two assume only rectangular meshes. This advantage is quite crucial when we are dealing with earth resistivity measurements since, as has been discussed, resistivity is sensitive to terrain anomalies. By using the FEM method the mesh can be adjusted to the local geomorphology, thus enabling the pseudo-noise due to terrain anomalies to be taken into account (Fox et al., 1980; Molano et al., 1990) It is this feature that renders the FEM superior to RNAM and FD methods, and therefore it will be used for this work.

As far as the implementation of the methods is concerned it can be argued that both RNAM and FDM are easier to develop since the mesh design principle is quite straightforward. The FEM, on the other hand, (for reasons that will be given in a following section) is a more laborious procedure as far as mesh construction is concerned.

3.1.3 Modelling dimensions

3-D Modelling - The resistivity forward problem is clearly a 3-D one: the measured AR is the effect of a three-dimensional current flow in to an earth which has a resistivity distribution varying in three dimensions. 3-D differential modelling schemes have been presented for the resistivity case: Dey and Morisson (1979b) produced a 3-D FD algorithm while Pridmore et al. (1981) presented a 3-D FEM algorithm.

[1]Recently, at least for the EM case, FD schemes which can incorporate curved boundaries have been presented (Taflove, 1995), however FEM still remains the most flexible method for modelling a variable topography.

The common feature of the 3-D differential schemes is that they are memory and time consuming. Although they offer the most accurate way to model the subsurface they are not widely used in inversion algorithms. This is because an iterative inversion algorithm needs the repetition of the forward model several (perhaps very many) times and hence an algorithm involving a 3-D modelling procedure will be extremely time consuming. Few examples of using 3-D schemes for inversion exist in the literature (Park and Van, 1991; Zhang et al., 1995). Recently, due to the advent of automatic measuring systems and the increase in computing power, the usage of 3-D differential models has increased, but they still remain a costly choice for routine data modelling and interpretation [2].

2-D Modelling- A far simpler approach to the resistivity forward problem is 2-D modelling: the measured signal is considered to be the effect of a 2-D current flow, while the subsurface resistivity is considered to vary in only two dimensions. Essentially, a 2-D model studies the behaviour of an idealized section of the subsurface. The 2-D schemes that have been produced (Muffti, 1976; Dittmer and Szymanski, 1993) can be only used for qualitative studies of the resistivity responses of the models since they use quite crude modelling assumptions. 2-D modelling schemes have the advantage of being economical in computer time and memory but their use in the accurate inversion of real data is inappropriate.

2.5-D Modelling- A compromise between these two approaches is a scheme known as modelling in two-and-a-half dimensions (2.5-D), where the change in resistivity is considered to be two dimensional but the current flow pattern is a three dimensional one. In other words, the measured values correspond to a three dimensional infinite half-space where the resistivity is allowed to vary in only two dimensions and remains constant in the strike direction.

The advantage of the 2.5-D approach is that a physically realistic representation,

[2]Further development of the computer systems will enable 3-D modelling to be widely used so there is no doubt that in few years time 3-D schemes will be the standard procedure for modelling and interpreting resistivity data.

involving full 3-D potential variations, is obtained by solving several problems with a restricted 2-D geometry. In this way the computational loads are reduced compared to a full 3-D scheme.

At the same time 2.5-D modelling provides relatively accurate results as long as the 2-D resistivity variation assumption is not strongly violated. In particular, Dey and Morisson (1979b) presented comparative dipole-dipole array results between their 2.5-D and 3-D FDM schemes and concluded that 2.5-D models become inaccurate only when the body extends along the strike direction a distance smaller than the maximum dipole separation. However, for many subsurface features the 2.5-D assumption is quite close to reality even if this criterion is violated and the consequent error introduced is not prohibitively large.

In conclusion, 2.5-D modelling is a good compromise between computing load and accuracy, and therefore it is used in this work.

3.1.4 General FEM principles

The FEM was initially developed for mechanical and civil engineering problems but soon found application in a wider range of problems (such as electromagnetic, magnetostatic, acoustic, heat and water conduction) and is still an active area of research (Zienkiewicz and Taylor, 1989). Here, a summary of the basic principles of the FEM is presented.

Trial Solution - The FEM method seeks an approximate solution of the governing differential equation in terms of an unknown function U which satisfies both the equation and the boundary conditions. The approximate solution U_a is obtained by using the classical trial-solution procedure which has the form of a finite sum of functions. The general form of the solution is:

$$U_a = \sum_{i=1}^{N} u_i \phi_i \tag{3.3}$$

The undetermined coefficients u_1 , u_2 , ..., u_N represent the values of the function at the N points which are defined when the domain is discretized.
The functions ϕ_0, ϕ_1 , ..., ϕ_N are called trial, or base, or shape functions. Most commonly, they are expressed as simple power series polynomials (their exact form depends on the way the domain is discretized).

Galerkin Method - As long as the functions ϕ_i have been specified, only the parameters u_i remain unknown. Each particular set of u_i values defines uniquely a solution for U_a. However, since i=1,2,...N there are N^2 unique solutions. Hence, a method of defining a set of u_i parameters so as to allow the "best" approximate solution U_a to be acquired should be used. By "best" is meant the approximate solution that is as close as possible to the true one.

If the general form of the field equation is $G(U) = F$, then by replacing U with the approximate solution we get $G(U_a) \simeq F$ since U_a is only approximately equal to U. Therefore, the residual caused by the approximation is:

$$R_a = G(U_a) - F \tag{3.4}$$

The formulation of the problem is now as follows: use an optimization criterion in order to obtain the set of u_i parameters for which the residual R_a of equation 3.4 becomes minimal. Several criteria have been suggested but the most popular in FE analysis is the Galerkin weighted residual method (GM) [3].

In the GM the residual R_a minimization is achieved by forcing the weighted average of R_a for each parameter u_i to be zero over the entire domain D. The weighted functions are the trial functions ϕ_i associated with each u_i:

$$\int_D R_a \, \phi_i(x,y) dD = 0 \qquad i = 1, 2..N \tag{3.5}$$

[3]The Ritz variational method is also widely used, and despite the fact that is based on different concepts, gives results identical to the GM when applied to identical problems (Pridmore et al.,1981). However, according to Burnett (1989) the GM is applicable to a wider class of problems and so is preferable.

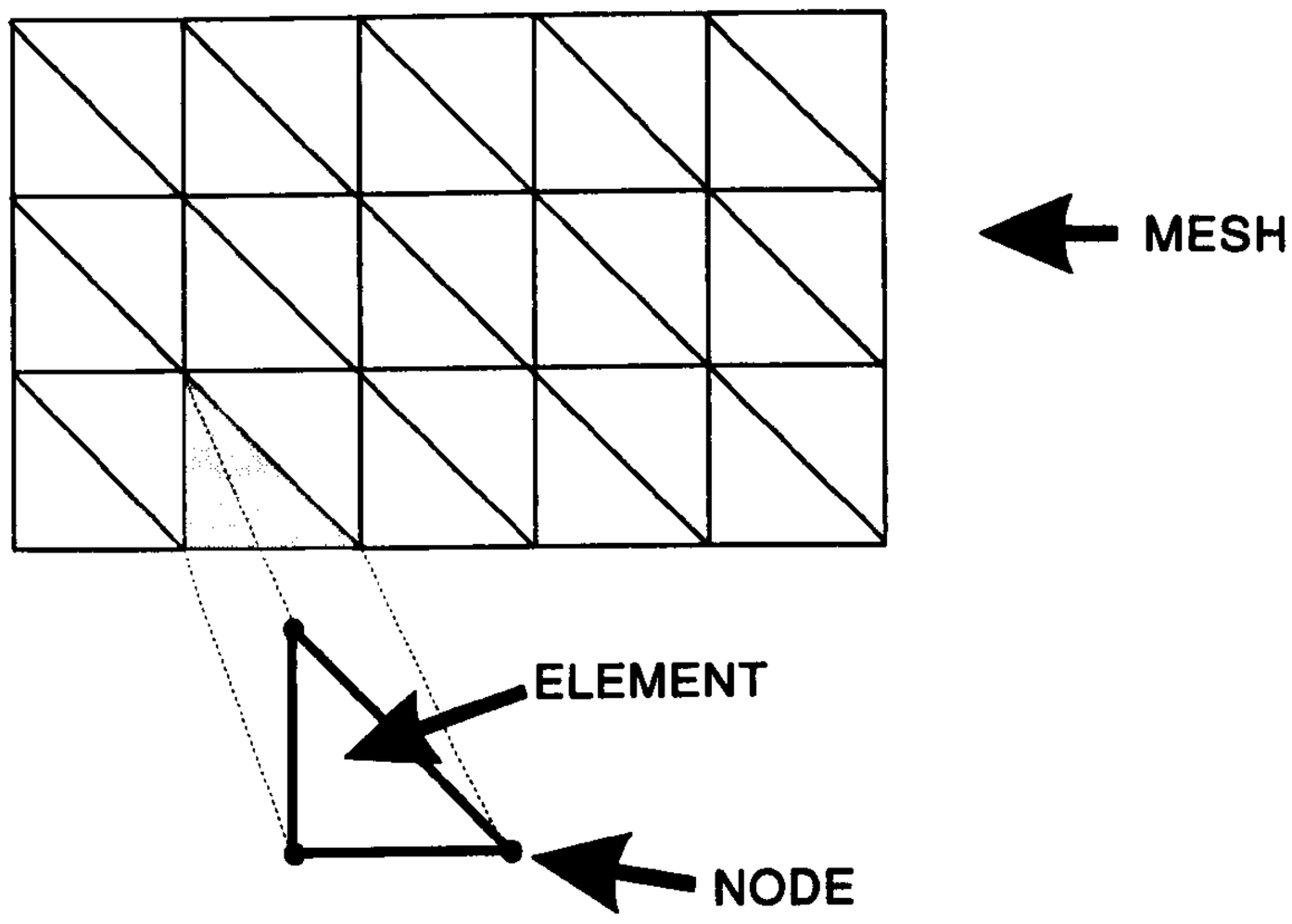


Figure 3.1: An example of FEM discretization.

N algebraic equations of this form are required, each using different trial functions, in order to determine the N parameters u_i. By using the GM two things are achieved:
a) The initial problem of seeking the solution of the unknown function $U(x, y)$ is changed to finding the unknown parameters u_i which are the values of the function at specific points.
b) the unsolvable initial equation is transformed to an approximately equivalent but potentially solvable system of linear equations (see equation 3.5).

Discretization of the Domain - As mentioned above, the application of the FEM is based on finding the solution of the unknown function at discrete points of the domain. So far, the approximate solution given by equation 3.5 is defined for the entire domain. Since the trial functions are in the form of quite simple power series, a solution for the entire domain would not be able to cope with any small or local changes of the unknown function or its gradients and hence the solution would be undesirably inaccurate. In order to handle this problem the domain is subdivided into a finite number of smaller regions (sub-domains) called elements. The elements have simple shapes (e.g. triangular or quadrilateral for the 2-D case) and they are assembled in a pattern known as a mesh. Each element has its own power series approximate solution (see equation 3.3) which is transformed into a series of linear

algebraic equations (element equations) using the GM [4].

The approximate solution is defined at particular points of the elements called nodes. Each type of element has a specific number and pattern of nodes and for elements of the same type the element equations are algebraically identical. This is extremely useful as the element equations need only be derived for a few typical elements of the mesh. In Figure 3.1 an example of a FEM mesh is shown.

Next, the element equations are assembled into a far larger set of algebraic equations Known as the system equations. It is at this stage that the boundary conditions must be applied. The solution of this system will give the unknown parameters u_i which describe the discrete solutions of the function at the nodes.

3.2 2.5-D FEM Resistivity Modelling

The first 2.5-D FEM algorithm for the resistivity case is due to the pioneering work of Coggon (1971). Rijo (1977) presented a modified 2.5-D FEM scheme and gave a detailed analysis of its features. Several other schemes were presented later (Queralt et al.,1991). The 2.5-D FEM algorithm developed in this work is partly based on a 2-D scheme produced by Dittmer (Dittmer and Szymanski, 1993) who modified a general 2-D program presented by Burnett (1989) and adjusted it to work for the 2-D resistivity case.

Although the previously mentioned researchers gave a detailed analysis of the application of the 2.5-D FEM in the resistivity modelling, the scheme developed for that work will be described in full. This is necessary for two reasons:

a) The scheme was developed from scratch and in many ways is different to the existing ones.

b) Full justification of techniques that are presented in this work (a strategy

[4]The only difference from the set of equations 3.5 is that the error minimization is taking place at each sub-domain and not over the entire domain

for modelling arrays parallel to the strike direction, study of the topographical effect and the calculation of the Jacobian matrix) is possible only if the core of the 2.5-D FEM modelling is fully explained.

3.2.1 Formulation of the field equations

As demonstrated in section 2.1.3 the governing equation for the steady state current electrical field when there are no charge sources or sinks in the medium is (equation 2.7):

$$\nabla.(-\sigma_{(x,y,z)}\nabla V_{(x,y,z)}) \;=\; 0$$

In the presence of a current source the previous equation no longer holds. If J is the current density then equation 2.6 becomes (see equation 2.5)

$$\nabla.(-\sigma_{(x,y,z)}\nabla V_{(x,y,z)}) \;=\; \nabla \cdot J \tag{3.6}$$

∇ is a three dimensional operator and the term $\nabla \cdot J$ describes the current source. In reality the current is applied via finite probes, however for the purposes of the modelling the current is considered to be introduced by point sources. Therefore, the source term $\nabla \cdot J$ can be described by a Dirac delta function and a point current I (Coggon, 1971). If $x_s,\ y_s,\ z_s$ are the source coordinates the source term is:

$$\nabla \cdot J \;=\; I\delta(x - x_s)\delta(y - y_s)\delta(z - z_s) \tag{3.7}$$

Taking in to account the 2.5-D modelling assumptions (3-D source, 2-D resistivity variation) the expression of the field equation will be

$$\nabla.(-\sigma_{(x,z)}\nabla V_{(x,y,z)}) \;=\; I\delta(x - x_s)\delta(y - y_s)\delta(z - z_s) \tag{3.8}$$

The potential field depends on all three coordinates and hence a way should be found to include this dependence within the FEM formulation. The classical way is to Fourier transform the potential variation in the y direction into the wavenumber domain (Coggon, 1971). Because of the constant resistivity in the strike direction it

follows that the potential $V_{(x,y,z)}$ is an even function of y so we can apply the cosine Fourier transform. The transformed potential is given by

$$\tilde{V}_{(x,k,z)} = \int_0^\infty V_{(x,y,z)} \cos(ky) dy \tag{3.9}$$

Therefore the field equation 3.8 when the transformed potential is considered takes the form

$$\nabla.(-\sigma_{(x,z)} \nabla \tilde{V}_{(x,k,z)}) = I\delta(x - x_s)\delta(z - z_s) \tag{3.10}$$

3.2.2 General form of element equations

In the FEM the domain is discretized into elements. Assuming that each element is constant and homogenous equation 3.10 becomes:

$$-\sigma_{(x,z)} \nabla^2 \tilde{V}_{(x,k,z)} = I\delta(x - x_s)\delta(z - z_s) \tag{3.11}$$

By using known cosine Fourier properties [5] equation 3.11 takes the form:

$$-\sigma_{(x,z)} \nabla^2 \tilde{V}_{(x,k,z)} + \sigma_{(x,z)} k^2 \tilde{V}_{(x,k,z)} = I\delta(x - x_s)\delta(z - z_s) \tag{3.12}$$

(the operator ∇ is now a 2-D one). The initial differential equation is now transformed into a Helmholtz type equation in the wavenumber domain. Based on equation 3.12, we can derive the general element equations via the procedure summarized in the previous section.

The trial solution approximating the transformed potential $\tilde{V}_{(x,k,z)}$ at the element e is

$$\tilde{V}_{(x,k,z)} = \sum_{i=1}^{N} \tilde{a}_i \, \phi_i(x, z) \tag{3.13}$$

where $\tilde{a}_i$ is the unknown transformed potential at the element's nodes and ϕ_i are the shape functions to be defined later.

[5] If U_k is the transformed function of U_y in respect of k it can be shown that for the second derivative of U_k it holds that (Kreyszig, 1992): $U_k'' = -k^2 U_k$

Following equation 3.5, the set of the Galerkin residual equations for the element can be derived [6]):

$$-\sigma \int\int_e \left[\frac{\partial}{\partial x}\left(\frac{\partial \tilde{V}}{\partial x}\right)\phi_i + \frac{\partial}{\partial z}\left(\frac{\partial \tilde{V}}{\partial z}\right)\phi_i \right] dxdz + \sigma \int\int_e k^2 \tilde{V}\phi_i \, dxdz$$
$$= \int\int_e f\phi_i dxdz \quad i = 1..N \tag{3.14}$$

At the next step the high order differential terms of equation 3.14

$$\int\int_e \frac{\partial}{\partial l}\left(\frac{\partial \tilde{V}}{\partial l}\right) \phi_i \, dxdz \quad i = 1..N \quad l = x, z$$

can be integrated by parts using the chain rule of differentiation, that is

$$\int\int_e \frac{\partial}{\partial l}\left(\frac{\partial \tilde{V}}{\partial l}\right) \phi_i \, dxdz = \int\int_e \left[\frac{\partial}{\partial l}\left(\frac{\partial \tilde{V}}{\partial l}\,\phi_i\right)\right] dxdz - \int\int_e \frac{\partial \tilde{V}}{\partial l}\,\frac{\partial \phi_i}{\partial l}\, dxdz \tag{3.15}$$

Conducting the appropriate substitutions in equation 3.14 according to the pattern of equation 3.15 and rearranging the terms yields

$$\sigma \int\int_e \left(\frac{\partial \tilde{V}}{\partial x}\,\frac{\partial \phi_i}{\partial x} + \frac{\partial \tilde{V}}{\partial z}\,\frac{\partial \phi_i}{\partial z} \right) dxdz + \sigma \int\int_e k^2 \tilde{V}\phi_i \, dxdz =$$
$$\int\int_e f\,\phi_i \, dxdz + \sigma \int\int_e \left[\frac{\partial}{\partial x}\left(\frac{\partial \tilde{V}}{\partial x}\,\phi_i\right) + \frac{\partial}{\partial z}\left(\frac{\partial \tilde{V}}{\partial z}\,\phi_i\right) \right] dxdz$$
$$i = 1..N \tag{3.16}$$

The last term of equation 3.16 is a perfect differential.

By using the 2-D divergence (Green's) theorem this term can be reduced to a line integral over the boundary of the element's area

$$\sigma \int\int_e \left[\frac{\partial}{\partial x}\left(\frac{\partial \tilde{V}}{\partial x}\,\phi_i\right) + \frac{\partial}{\partial z}\left(\frac{\partial \tilde{V}}{\partial z}\,\phi_i\right) \right] dxdz =$$
$$\oint_e \left[\sigma\frac{\partial \tilde{V}}{\partial x}\,\phi_i\, n_x + \sigma\frac{\partial \tilde{V}}{\partial z}\,\phi_i\, n_z \right] ds = \oint_e T_n\phi_i \, ds \tag{3.17}$$

---

[6] A simplified notation is used:

Transformed potential: $\tilde{V}_{(x,k,z)} \rightarrow \tilde{V}$

2-D Conductivity: $\sigma_{(x,z)} \rightarrow \sigma$

Source term: $I\delta(x - x_s)\delta(z - z_s) \rightarrow f$.

The n_x, n_z symbols are the x,z components of the outward unit n normal to the element boundary (the inward unit normal to the boundary is -n). The term T_n can describe the energy flowing out from the system (the inward flowing energy is T_{-n}). The main reason for integrating by parts was to create this term which contains the first derivative of the potential. As will be shown later, this term facilitates the application of the boundary conditions. Equation 3.16 now becomes

$$\sigma \int\int_e \left(\frac{\partial \tilde{V}}{\partial x} \frac{\partial \phi_i}{\partial x} + \frac{\partial \tilde{V}}{\partial z} \frac{\partial \phi_i}{\partial z} \right) dxdz + \sigma \int\int_e k^2 \tilde{V} \phi_i \, dxdz =$$

$$\int\int_e f \, \phi_i \, dxdz - \oint_e T_n \phi_i \, ds \qquad i = 1..N \tag{3.18}$$

Now, the general form of the trial solution (equation 3.13) is substituted into equation 3.18 yielding

$$\sum_{j=1}^{N} \sigma \left[\int\int_e \frac{\partial \phi j}{\partial x} \frac{\partial \phi_i}{\partial x} dxdz + \int\int_e \frac{\partial \phi_j}{\partial z} \frac{\partial \phi_i}{\partial z} dxdz + \int\int_e k^2 \phi_j \phi_i dxdz \right] \tilde{a}_j$$

$$= \int\int_e f \, \phi_i \, dxdz - \oint_e T_n \phi_i \, ds \qquad i = 1..N \tag{3.19}$$

These are the element equations for a typical element. If

$$K_{ij}^e = \sigma \int\int_e \frac{\partial \phi j}{\partial x} \frac{\partial \phi i}{\partial x} dxdz + \sigma \int\int_e \frac{\partial \phi j}{\partial z} \frac{\partial \phi i}{\partial z} dxdz + \int\int_e k^2 \phi_{k\,j} \phi_{k\,i} dxdz \tag{3.20}$$

and

$$F_i^e = \int\int_e f \phi_i dxdz - \oint_e T_n \phi_i ds \tag{3.21}$$

the element equations can be written in a matrix form giving

$$\begin{bmatrix} K_{11}^e & K_{12}^e & \dots & K_{1N}^e \\ K_{21}^e & K_{22}^e & \dots & K_{2N}^e \\ \vdots & \vdots & & \vdots \\ K_{N1}^e & K_{N2}^e & \dots & K_{NN}^e \end{bmatrix} \begin{bmatrix} \tilde{a}_1^e \\ \tilde{a}_2^e \\ \vdots \\ \tilde{a}_N^e \end{bmatrix} = \begin{bmatrix} F_1^e \\ F_2^e \\ \vdots \\ F_N^e \end{bmatrix} \tag{3.22}$$

which in a simple matrix notation is $\mathbf{K}^{(e)}\ \mathbf{A}^{(e)} = \mathbf{F}^{(e)}$. The terms K_{ij}^e are called stiffness terms and the matrix $\mathbf{K}^{(e)}$ is the element stiffness matrix [7] while the terms $f_i^{(e)}$ are the load terms.

[7] The name is due to the initial use of the FEM in structural civil engineering problems

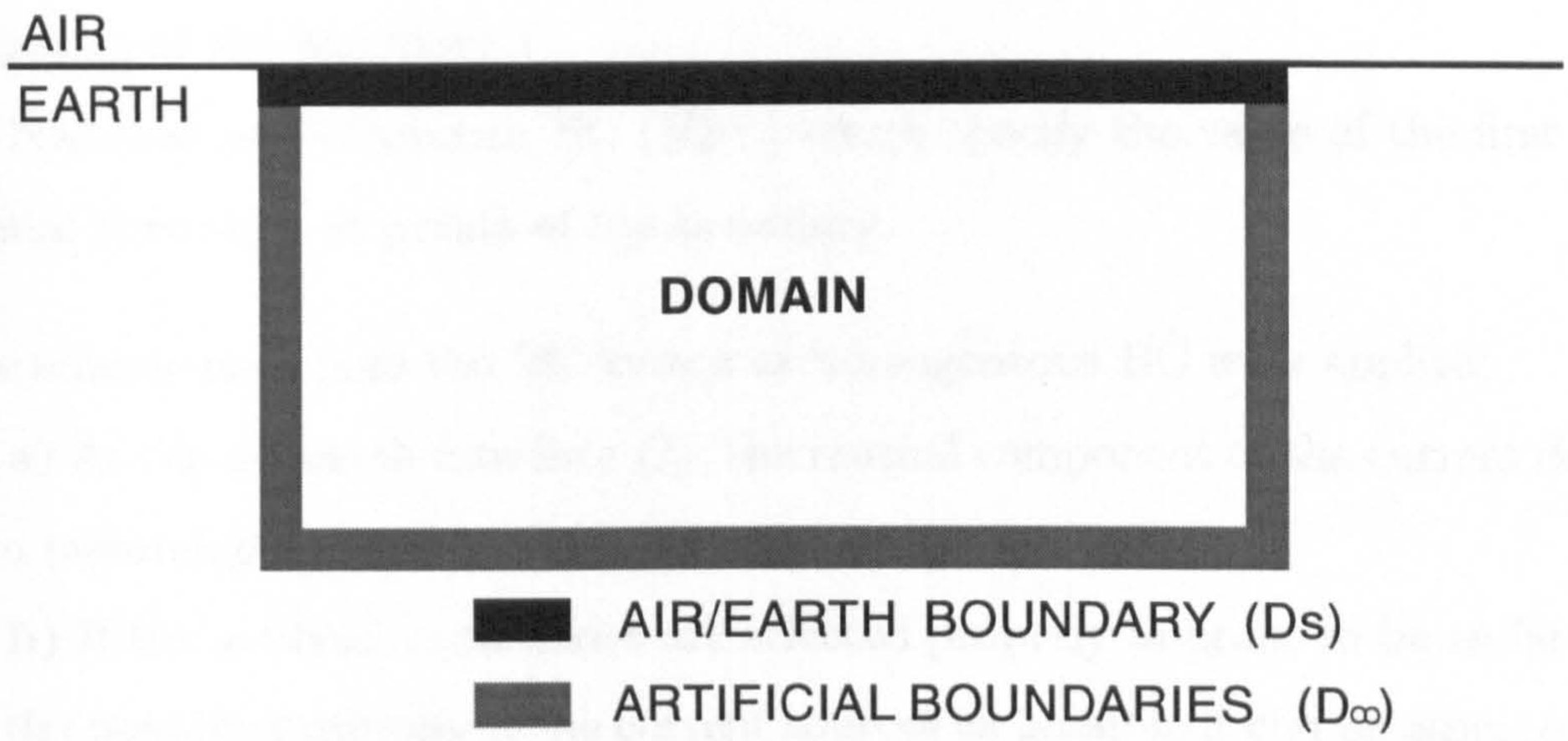


Figure 3.2: The boundaries of the domain.

The element equations, which so far have been developed in a general form, will become meaningful as long as the trial solution is defined (this is closely related to the type of elements that exist in the mesh).

3.2.3 Boundary conditions

In order to obtain a unique solution the boundary conditions have to be considered. In our case there are two types of boundary conditions: domain boundary conditions, or simply boundary conditions (BC), and inter-element boundary conditions (IBC).

Domain BC -: These are equations that must be satisfied by the exact solutions on the boundaries of the domain (D). In our problem the domain is enclosed by a boundary that is formed of two parts: a natural boundary, D_s, which lies at the Air/Subsurface interface and an artificial one, D_∞, which simulates the potential behaviour at an infinite distance from the source both in the horizontal and vertical directions (see Figure 3.2).

For this boundary-value problem, which involves the solution of a second order differential equation, two types of BC can be applied.

Essential or Dirichlet BC (EBC) which specify the value of the potential

U at points of the boundary.

Natural or Neumann BC (NBC) which specify the value of the first order potential derivative at points of the boundary.

In the scheme used here the BC known as homogeneous BC were applied:

a) At the air/earth interface D_S, the normal component of the current density is zero (assuming that air has infinite resistance).

b) If the artificial boundaries are selected properly in order to be as far away from the possible positions of the current sources as possible it can be assumed that the potential at the boundary D_∞ is zero ($V_{D_\infty} = 0$).

Inter-element BC - These boundary conditions must be satisfied by the exact solution on the boundaries between elements. For the resistivity case these conditions are the continuity of the potential, and the continuity of the normal component of the current density vector between the boundaries. Hence, if two elements (e), (f), with conductivity σ_e, σ_f respectively have a common boundary IB and **n** is the unit normal to that boundary, for every point at that boundary

$$\begin{aligned} V^e_{IB} &= V^f_{IB} \\ \sigma_e \frac{\partial V^e_{IB}}{\partial n} &= \sigma_f \frac{\partial V^f_{IB}}{\partial n} \end{aligned} \tag{3.23}$$

Note that any possible trial solution must be selected so as to satisfy these IBC.

3.2.4 Trial solution for the linear triangular element

There are several type of elements, having a pattern of nodes of varying complexity; however, for this work, only simple linear elements of triangular shape were used (three nodes, one at each vertex). This is partly in order to avoid complexity and partly because triangles can simulate almost any structure met in geophysical exploration with reasonable accuracy. That is why this element type is so popular in geophysical modelling.

Suppose that the general form of the trial solution of the element is given by a simple

power series approximation (see Figure 3.3a)

$$\tilde{V} = b + cx + dz \tag{3.24}$$

Instead of expressing the trial solution in respect of the parameters b,c,d it is preferable to express it in respect of the values of the trial solution at the nodes in the form of interpolation polynomials since in this way we can ensure more easily the continuity of the potential at the inter-element boundaries.

In particular, consider the two elements (e), (f) which have two common nodes (see Figure 3.3b). Since for both elements the trial solutions $\tilde{V}_e$, $\tilde{V}_f$ are linear polynomials, along the elements' sides both $\tilde{V}_e$ and $\tilde{V}_f$ will be straight lines. The continuity between the e, f common boundary will be achieved only if these lines are identical. A straightforward way of accomplishing this is to force the trial solutions of the two elements to be identical at their common nodes (a straight line is uniquely determined by two points), a thing that requires the element's trial solution to be expressed in terms of its values at the nodes.

This can be done in the following way: let 1,2,3 be the nodes of the element , (x_1, z_1), (x_2, z_2), (x_3, z_3) their respective coordinates and $\tilde{a}_1$, $\tilde{a}_2$, $\tilde{a}_3$ their respective trial solutions. By using equation 3.24

$$\begin{aligned} b + cx_1 + dz_1 &= \tilde{a}_1 \\ b + cx_2 + dz_2 &= \tilde{a}_2 \\ b + cx_3 + dz_3 &= \tilde{a}_3 \end{aligned} \tag{3.25}$$

Solving equation 3.25 for b,c,d, substituting the resulting expressions into equation 3.24 and rearranging the terms gives the interpolatory form of $\tilde{V}$

$$\tilde{V} = \sum_{j=1}^{3} \tilde{a}_j \, \phi_j(x, z) \tag{3.26}$$

where

$$\phi_j = \frac{A_j + B_j x + C_j z}{2\Delta} \qquad j = 1, 2, 3 \tag{3.27}$$

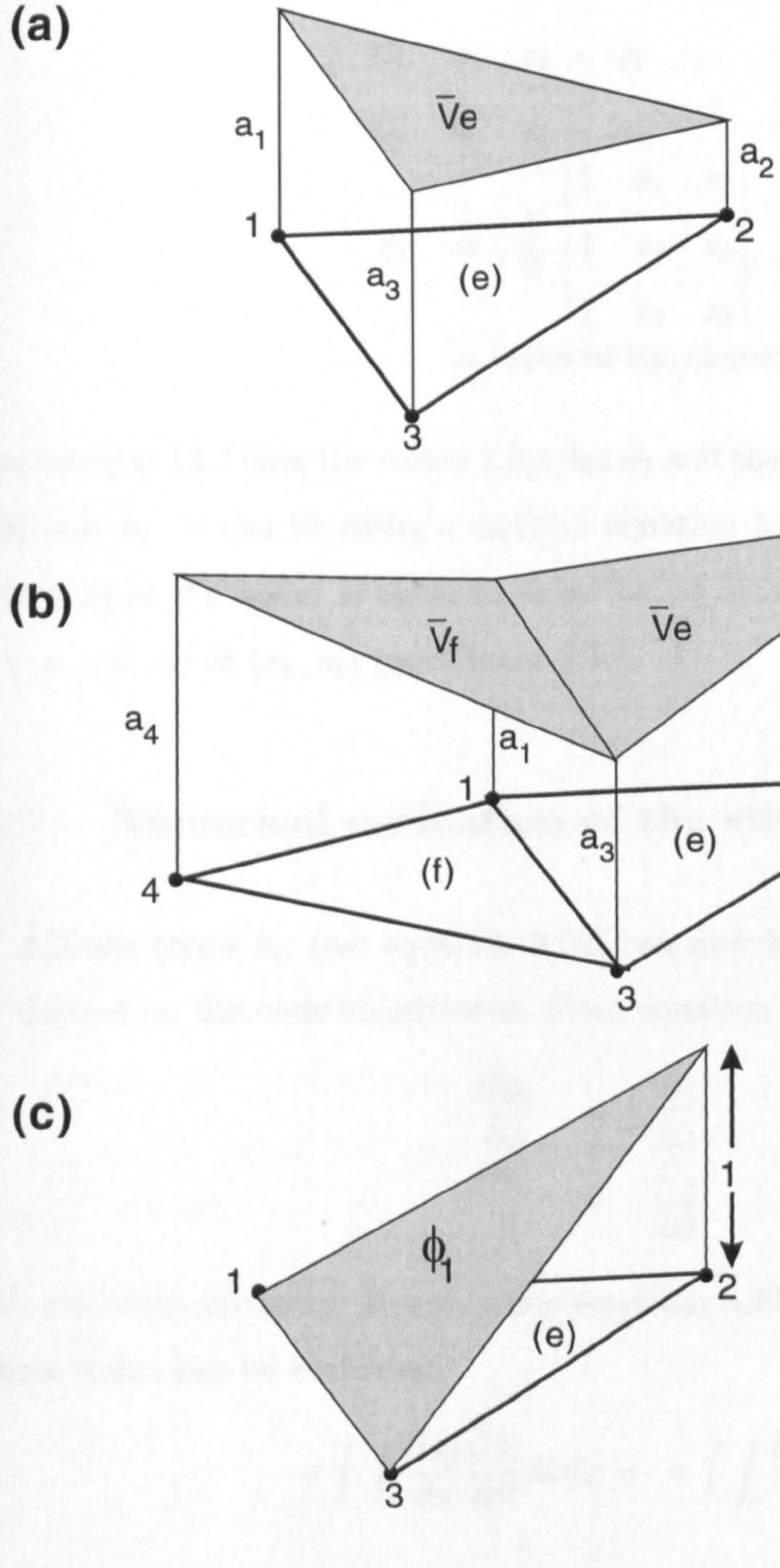


Figure 3.3: a) The trial solution of one element, b) the trial solutions of two elements, c) the trial function of node 2.

and

$$\begin{aligned} A_j &= x_k z_l - x_l z_k \\ B_j &= z_k - z_l \\ C_j &= x_l - x_k \\ \Delta &= \frac{1}{2}\begin{vmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ 1 & x_3 & z_3 \end{vmatrix} \\ &= \text{area of the element} \end{aligned} \tag{3.28}$$

The subscripts j,k,l have the values 1,2,3, for ϕ_1 and they change their order cyclically for ϕ_2 and ϕ_3. It can be easily seen from equation 3.27 that the value of the trial function ϕ_j at the nodes is equal to unity for $(x, z) = (x_j, z_j)$ while it is zero for $(x, z) = (x_l, z_l)$ or (x_k, z_k) (see Figure 3.3c).

3.2.5 Numerical evaluation of the stiffness and load terms

The stiffness terms k_{ij} (see equation 3.20) can now be evaluated easily since they only depend on the node coordinates. From equation 3.27

$$\begin{aligned} \frac{\partial \phi_j}{\partial x} &= \frac{B_j}{2\Delta} \\ \frac{\partial \phi_j}{\partial y} &= \frac{C_j}{2\Delta} \end{aligned} \tag{3.29}$$

which are both constants. Hence, using equation 3.29 the first two integrals of the stiffness terms can be evaluated.

$$\begin{aligned} & \sigma \int\int_e \frac{\partial \phi_j}{\partial x}\frac{\partial \phi_i}{\partial x} dx dz + \sigma \int\int_e \frac{\partial \phi_j}{\partial z}\frac{\partial \phi_i}{\partial z} dx dz \\ &= \sigma \frac{B_i B_j}{4\Delta^2} \int\int_e dx dz + \sigma \frac{C_i C_j}{4\Delta^2} \int\int_e dx dz \\ &= \frac{\sigma}{4\Delta}(B_i B_j + C_i C_j) \quad i, j = 1, 2, 3 \end{aligned} \tag{3.30}$$

the remaining integral of the stiffness terms

$$\sigma \int \int_e k^2 \phi_j \phi_i dx dz \qquad i,j \; = \; 1,2,3 \tag{3.31}$$

is evaluated by using a triangle integration formula widely employed in FEM analysis (Rao, 1985). This formula involves the area (or barycentric) coordinates of a triangle which are three local coordinates related to the triangular geometry. If f_1, f_2, f_3 are the area coordinates of the triangle $\widehat{123}$ with area Δ, raised to the powers l,m,n then

$$\int \int_e f_1^l f_2^m, f_3^n \; dxdz \; = \; \frac{l!m!n!}{(l+m+n+2)!} 2\Delta \tag{3.32}$$

For the linear triangle the area coordinates are identical to the trial functions (Burnett, 1989) and hence this integration formula can be applied directly to the numerical evaluation of the integral of equation 3.31 :

$$\sigma \int \int_e k^2 \phi_j \phi_i \; a_j dx dz = \begin{cases} k^2 \frac{\Delta}{6} & i = j \\ k^2 \frac{\Delta}{12} & i \neq j \end{cases} \tag{3.33}$$

The loading terms consist of two integrals. The first represents the current flowing into the element.

$$\int \int_e f \; \phi_i \; dxdz = \int \int_e I \delta(x - x_s) \delta(z - z_s) \phi_i \; dxdz \qquad i = 1,2,3 \tag{3.34}$$

If it is assumed that the current source is identical to a node, thus the current load must be assigned at the node and not at the whole element. If, for example, current is applied at node 1 then (as explained in the previous subsection) the value of the shape function ϕ_1 at that node is equal to 1 ($\phi_1 = 1$), while is 0 at the two remaining nodes ($\phi_2 = \phi_3 = 0$). Therefore, equation 3.34 will be (if x_1, z_1 are the coordinates of node 1)

$$\int \int_e f \phi_1 dx dz = \int \int_e I \delta(x - x_1) \delta(z - z_1) dx dz = I$$

$$\int \int_e f \phi_2 dx dz = 0$$

$$\int \int_e f \phi_2 dx dz = 0$$

Since the current is applied to a node, and nodes are being shared by more than one element, the amplitude of the current has to be equally shared among these

elements. If θ is the angle of the element's vertex which coincides with the source the correct magnitude I_s is given by (Rijo, 1977):

$$I_s = I\frac{\theta}{360} \tag{3.35}$$

Therefore the general form of the integral of equation 3.34 is:

$$\int\int_e I\delta(x-x_s)\delta(z-z_s)\phi_i dxdz = \begin{cases} I_s & i = s \\ 0 & i \neq s \qquad i = 1,2,3 \end{cases} \tag{3.36}$$

where s is the nodal source.

The remaining integral

$$\oint_e T_n\phi_i \, ds$$

represents expressions of the gradient of the potential at the element boundaries. These expressions need not to be calculated since they will be assigned values either when applying the boundary conditions or when assembling the system equations.

Taking the previous considerations into account, we can write the element equation for the 2-D resistivity problem

$$\left(\frac{\sigma}{4\Delta}\begin{bmatrix} B_iB_i+C_iC_i & B_iB_j+C_iC_j & B_iB_k+C_iC_k \\ B_jB_i+C_jC_i & B_jB_j+C_jC_j & B_jB_k+C_jC_k \\ B_kB_i+C_kC_i & B_kB_j+C_kC_j & B_kB_k+C_kC_k \end{bmatrix} + \frac{\sigma k^2\Delta}{12}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}\right)\cdot\begin{bmatrix} \tilde{a}_i \\ \tilde{a}_j \\ \tilde{a}_k \end{bmatrix} = I_s\begin{bmatrix} \delta_i+\beta_i \\ \delta_j+\beta_j \\ \delta_k+\beta_k \end{bmatrix} \tag{3.37}$$

where the indices i, j, k indicate the nodes of the element, B, C are expressions of the element's coordinates (equation 3.28), Δ is the area of the element, I_s is the current source distributed into the element, δ_i denotes unity when the node i coincides with the origin, and otherwise is zero (δ_j, δ_k take analogous expressions for nodes j,k respectively), and $\beta_{i,j,k}$ are the boundary terms to be evaluated later.

3.2.6 Global system

As long as the stiffness matrix has been numerically evaluated for every element, the element equations should be assembled into a single set of linear equations, taking into account the fact that elements will share common nodes and the trial solution has been designed to be identical for the common elements' nodes.

If the mesh has M elements, then M sets of element equations should be assembled. The pattern of the assembly can be easily derived (Burnet, 1989). The stiffness term $K_{ij}^{(e)}$ of element e will be added to the term K_{ij} of the global stiffness matrix $\mathbf{K}$. Similarly, the load term $F_i^{(e)}$ will be added to the term F_i of the global load vector $\mathbf{F}$. The global system will have the form

$$\mathbf{K} \cdot \mathbf{A} = \mathbf{F} \tag{3.38}$$

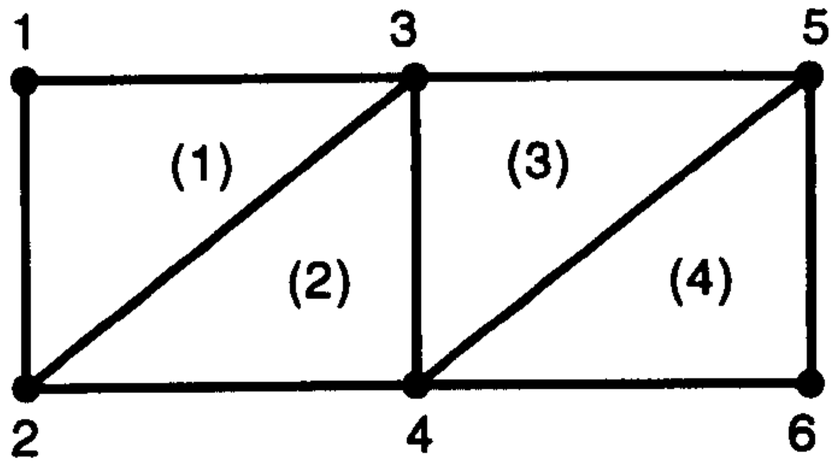
If the mesh has N nodes the matrix $\mathbf{K}$ will have NxN dimensions and will be sparse, symmetrical and banded since only directly connected nodes will share a line. The transformed potential ($\mathbf{A}$) and the load ($\mathbf{F}$) are vectors with dimension N.

In Figure 3.4 the general form of the global system of equations can be seen for an example with 4 elements and 6 nodes. Note the symmetrical and banded nature of the global stiffness matrix.

3.2.7 Application of the BC, IBC

After the system is assembled both BC and IBC are applied. Numerically, the application of the NBC is straightforward. If one side of the triangular element is at the boundary $\mathbf{D_S}$ the respective term $\partial U/\partial z$ of equation 3.17 can be made zero. The EBC are applied at elements that have sides at the boundary $\mathbf{D}_\infty$ by assigning the zero value at the node's trial solution **a** . In this way they are applied, indirectly, as a constraint to the system equations [8].

[8]Note that the EBC could have been applied directly at the element equations if a further integration by parts of the first term of equation 3.18 had taken place.



$$\begin{bmatrix} K_{11}^{(1)} & K_{12}^{(1)} & K_{13}^{(1)} & 0 & 0 & 0 \\ K_{21}^{(1)} & K_{22}^{(1)}+K_{22}^{(2)} & K_{23}^{(1)}+K_{23}^{(2)} & K_{24}^{(2)} & 0 & 0 \\ K_{31}^{(1)} & K_{32}^{(1)}+K_{32}^{(2)} & K_{33}^{(1)}+K_{33}^{(2)}+K_{33}^{(3)} & K_{34}^{(2)}+K_{34}^{(3)} & K_{35}^{(3)} & 0 \\ 0 & K_{42}^{(2)} & K_{43}^{(2)}+K_{43}^{(3)} & K_{44}^{(2)}+K_{44}^{(3)}+K_{44}^{(4)} & K_{45}^{(3)}+K_{45}^{(4)} & K_{46}^{(4)} \\ 0 & 0 & K_{53}^{(3)} & K_{54}^{(3)}+K_{54}^{(4)} & K_{55}^{(3)}+K_{55}^{(4)} & K_{56}^{(4)} \\ 0 & 0 & 0 & K_{64}^{(4)} & K_{65}^{(4)} & K_{66}^{(4)} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} = \begin{bmatrix} f_1^{(1)} \\ f_2^{(1)}+f_2^{(2)} \\ f_3^{(1)}+f_3^{(2)}+f_3^{(3)} \\ f_4^{(2)}+f_4^{(3)}+f_4^{(4)} \\ f_5^{(3)}+f_5^{(4)} \\ f_6^{(4)} \end{bmatrix}$$

Figure 3.4: The FEM stiffness matrix for 4 elements.

Finally, the IBC are applied to the assembled terms of the **F** vector, since after the assembly the line integrals of the normal current density terms (see equation 3.17) of, say, two elements which share the same boundary, will appear as a sum at a line of the vector **F**. Since they are equal (equation 3.23) and of opposite sign their addition will give zero.

3.2.8 Derivation of the potential

After the application of the BC the resulting system can be solved with the help of standard methods that are known to perform better when sparse banded systems are involved (Gauss elimination in this work). This solution will give the transformed potential $\tilde{V}_{(x,k,z)}$ for specified values of k.

In order to obtain the potential $V_{(x,y,z,)}$ the inverse Fourier cosine transform should

be applied, that is

$$V_{(x,y,z)} = \frac{2}{\pi}\int_0^{\infty} \tilde{V}_{(x,k,z)} \cos(k\,y)dk \tag{3.39}$$

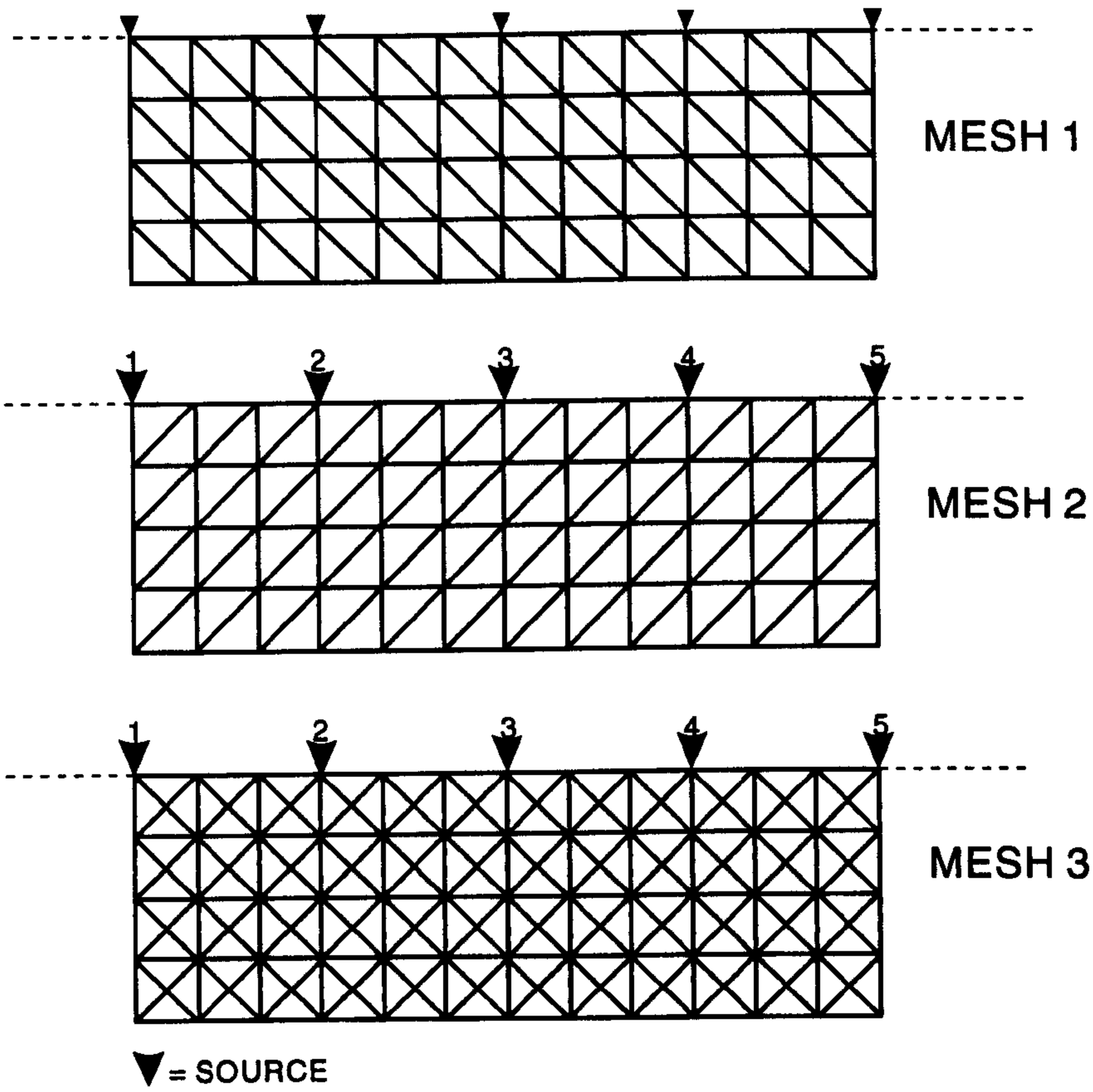
Hence, if the transformed potential $\tilde{V}_{(x,k,z)}$ is calculated for several k, the potential $V_{(x,y,z)}$ can be obtained by conducting numerical integration according to equation 3.39.

3.3 Practical Considerations and Application

3.3.1 Mesh design

The mesh design is one of the most important procedures in the FEM. The usual strategy in simulating soil is to separate the mesh into two main regions. In the region of main interest, the mesh consists of elements of approximately the same size repeated in regular patterns with a high nodal density. The area of interest is the one where the most accurate and valid solutions are obtained. This area occupies the central part of the domain and is surrounded by the second region in which the element size becomes gradually larger while the node density decreases. In this way the infinite distance from the source can be simulated. The practical considerations regarding the design of the mesh are:

a) The existence of a point source (which, as explained, is an artificial concept - no point sources exist in nature) imposes a singularity at the FEM calculations, in the sense that the potential at the source node is theoretically infinite. As long as no arrays parallel to the strike directions are modelled, there is no need to use the value of the transformed potential at that point. However, the singularity indicates abrupt potential variation (near logarithmic) at points near to the source. In addition the elements used (linear triangular) approximate the variation of the potential using linear polynomials which cannot cope with that such a variation in the potential. These factors introduce unavoidable inaccuracies into FEM solution.



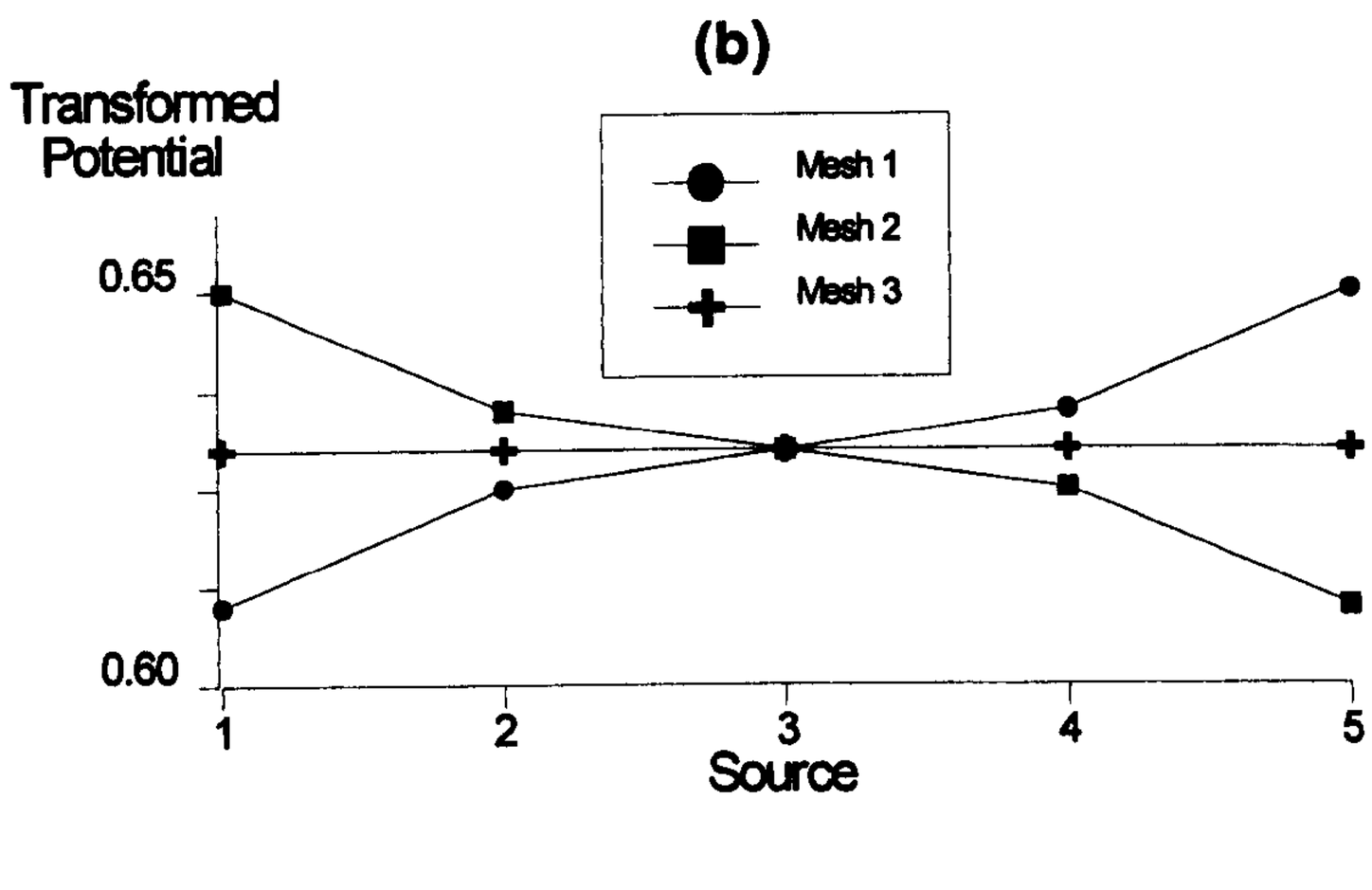


Figure 3.5: a) Three meshes with different element arrangement. b) The transformed potential calculated using those three meshes.

One solution [9] to this problem is to increase the refinement of the mesh in the area near to the probes in both the lateral and vertical directions. In practice, for the scheme used in this work it was found that there should be at least 2 nodes between the source node and the measuring node in order to achieve an acceptable accuracy. Greater refinement will further increase the accuracy, but there are computing cost limitations. Moreover, excessive refinement can lead to increasing ill-conditioning (Burnett, 1989).

b) Another important consideration in the design of a mesh is that the transition from refined to less refined areas should always be made in the smoothest possible manner. At the same time, if homogeneous boundary conditions are used, the boundaries have to extend a "great" distance from the possible source positions in order to simulate theoretical infinity (zero potential at the bottom and side boundaries).

In Figure 3.5a three meshes with different triangular element arrangements are shown. These meshes were used to calculate the transformed potential (for k=0.5) at a unit distance for every source position (also shown in Figure 3.5a). Since a homogeneous ground was considered the results for every source position should be identical. However, in Figure 3.5b it can be seen that only Mesh 3 produced the expected results. Also note that the results of Mesh 1 and Mesh 2 yield the results of Mesh 3 if they are averaged out. Therefore the element arrangement used in Mesh 3 (it can be described as a crossed triangular element arrangement) was followed for all of the meshes presented in this work. In Figure 3.6 a typical mesh used is depicted.

3.3.2 Solution of the global system of equations

The stiffness matrix of the global system of equations as explained is sparse, banded and symmetrical around the main diagonal. Because of these properties there is

[9] In a subsequent section another strategy for avoiding the singularity will be presented

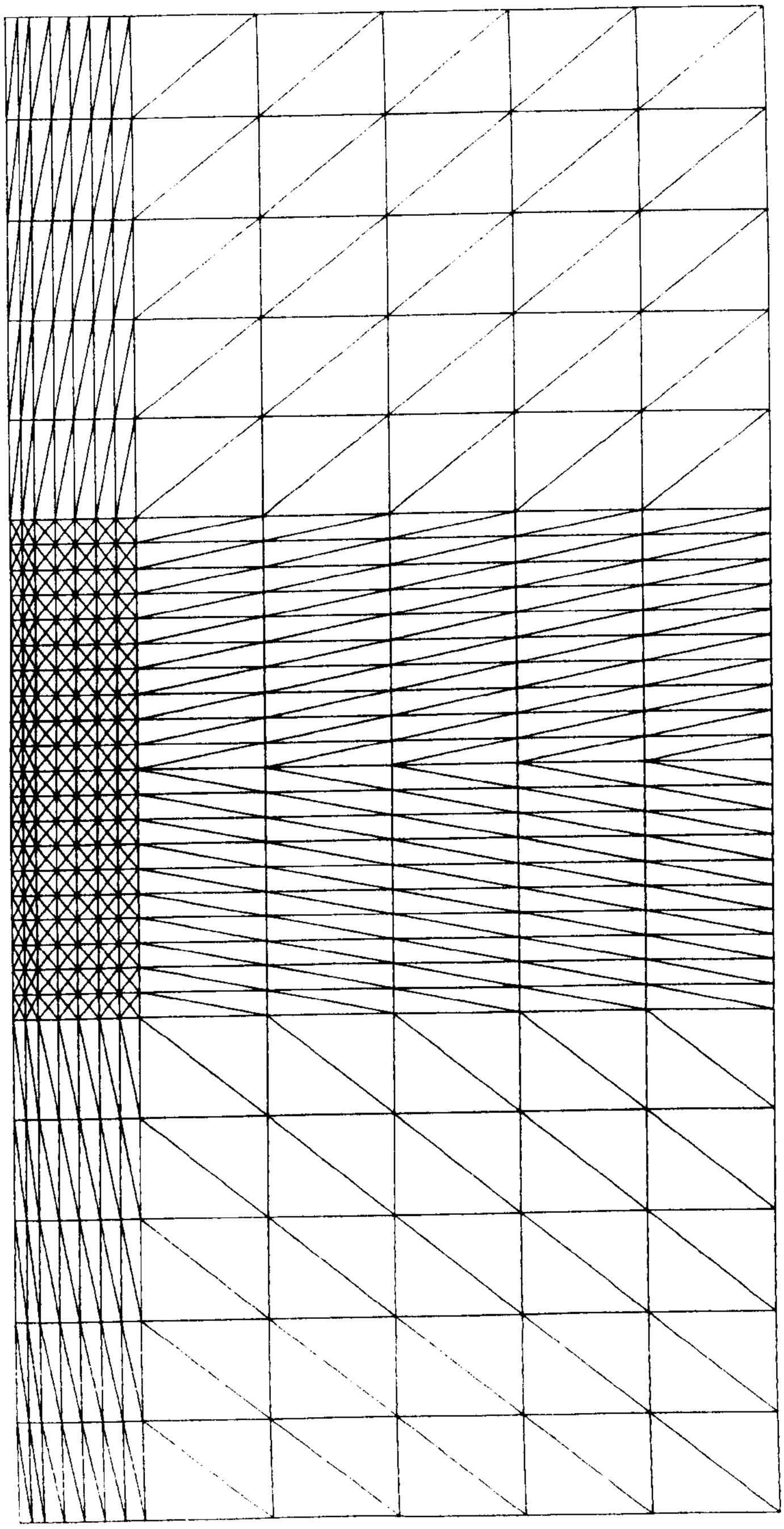


Figure 3.6: Typical mesh used in this work.

no need to store the entire matrix when solving the global system of equations. In particular, only the diagonal and the upper (or lower) elements of the matrix need to be stored, saving computer memory. If the system has N nodes the **K** matrix instead of being N x N is now N x HB where HB is the half-bandwith of the matrix (see Figure 3.7a). It can easily be seen that the half-bandwidth of the matrix is related to the nodal numbering. In particular if the mesh has M elements and N nodes:

$$HB = max[(N_{max}^{(e)} - N_{min}^{(e)}) + 1)] \qquad e = 1, 2, ...M$$

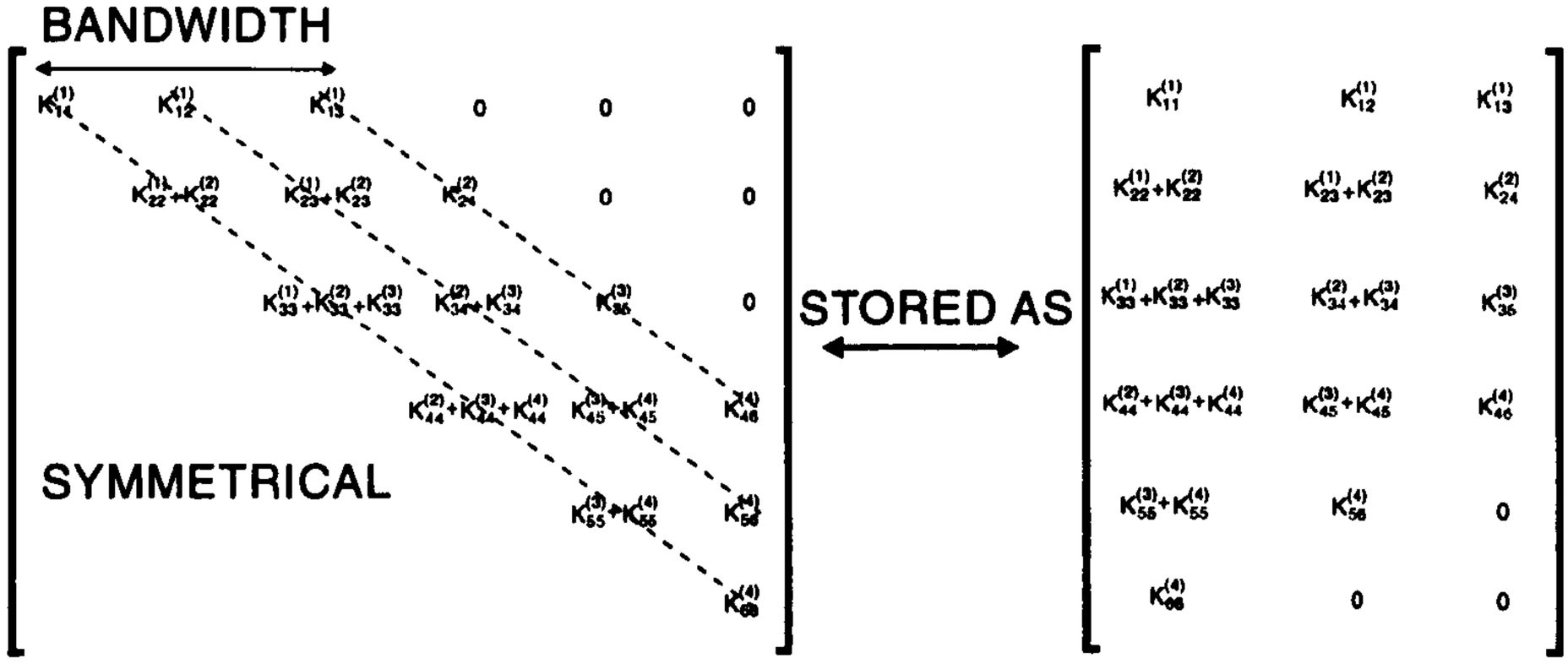
where $N_{max}^{(e)}$ and $N_{min}^{(e)}$ are the largest and smallest node numbers for element e. In Figure 3.7b it can be seen that different numbering of the same mesh results in different bandwidths. An optimum numbering will achieve the minimum half-bandwidth and therefore minimum storage cost.

The most popular techniques for solving this type of equations are Gauss elimination and Cholesky decomposition. In this work Gauss elimination was used[10]. Initially, it involves the forward reduction (elimination) of the linear system by a series of row and column operations; The solution is then given by back substitution.

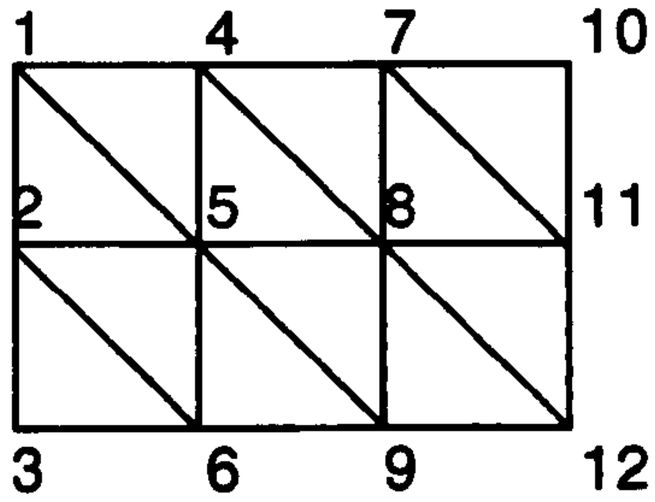
Note, at this point, that since resistivity modelling is being used to provide us with a series of simulated measurements (measured with the procedures described in Chapter 2), in order to obtain a full data set for the same resistivity distribution, the FEM numerical procedure has to be repeated for as many times as there are different sources (current positions). Fortunately, there is no need to repeat the entire procedure. Since the resistivity distribution remains constant, the matrix **K** will be the same and so its forward reduction will take place only once. Each additional source position will require only forward elimination of the global load vector **F** and back substitution in order to obtain the transformed potential.

[10]The routine used is specially designed to deal with sparce linear systems and is due to Burnett (1987)

(a)



(b)



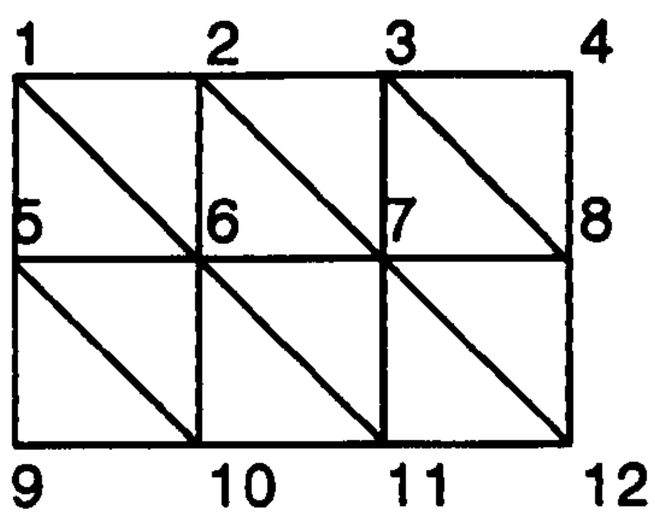


Figure 3.7: a) An example of how the stiffness matrix is actually stored. b) Different bandwidths resulting from different node numbering schemes.

3.3.3 Inverse Fourier transformation

In the 2.5-D FEM the transformed potential $\tilde{V}_{(x,k,z)}$ is calculated for several wavenumbers and the potential $V_{(x,y,z)}$ can be obtained by conducting numerical integration according to equation 3.39.

By the use of integral forms of Bessel functions (Abramowitz and Stegun, 1972) the analytical expression of the transformed potential for a homogeneous half-space can be derived:

$$\tilde{V} = \frac{I}{2\pi\sigma^P} B_0(k\,r) \tag{3.40}$$

Where B_0 is the modified Bessel function of zero order and r is the radial distance from the source. It is important to notice that the transformed potential is singular for $k\,r = 0$. When four electrode arrays are considered this singularity does not affect the accuracy of the integration, since it introduces errors which are cancelled out. For-two probe arrays, however, this singularity can become a serious source of errors. A method for tackling these errors is discussed in the following subsection.

In Figure 3.8 the spectrum of the potential for different distances from the point source for a homogeneous half-space is shown. We notice that the potential drops almost exponentially as the wavenumber is increased and has a quite smooth pattern.

For linear arrays, the cosine term in equation 3.39 disappears and this smooth pattern remains: the inverse Fourier transformation takes the form (y is considered to be zero at the measuring line):

$$V_{(x,0,z)} = \frac{2}{\pi}\int_0^\infty \tilde{V}_{(x,k,z)}dk \tag{3.41}$$

Simple integration strategies can be used in this case. One common way of calculating the area of the spectrum is by approximating it with a subsectional exponential fit and then integrating analytically (e.g. Dey and Morisson, 1979a).

How many k points should be used within an appropriate integration scheme? In general, increasing the number of k points makes the result more accurate but, on

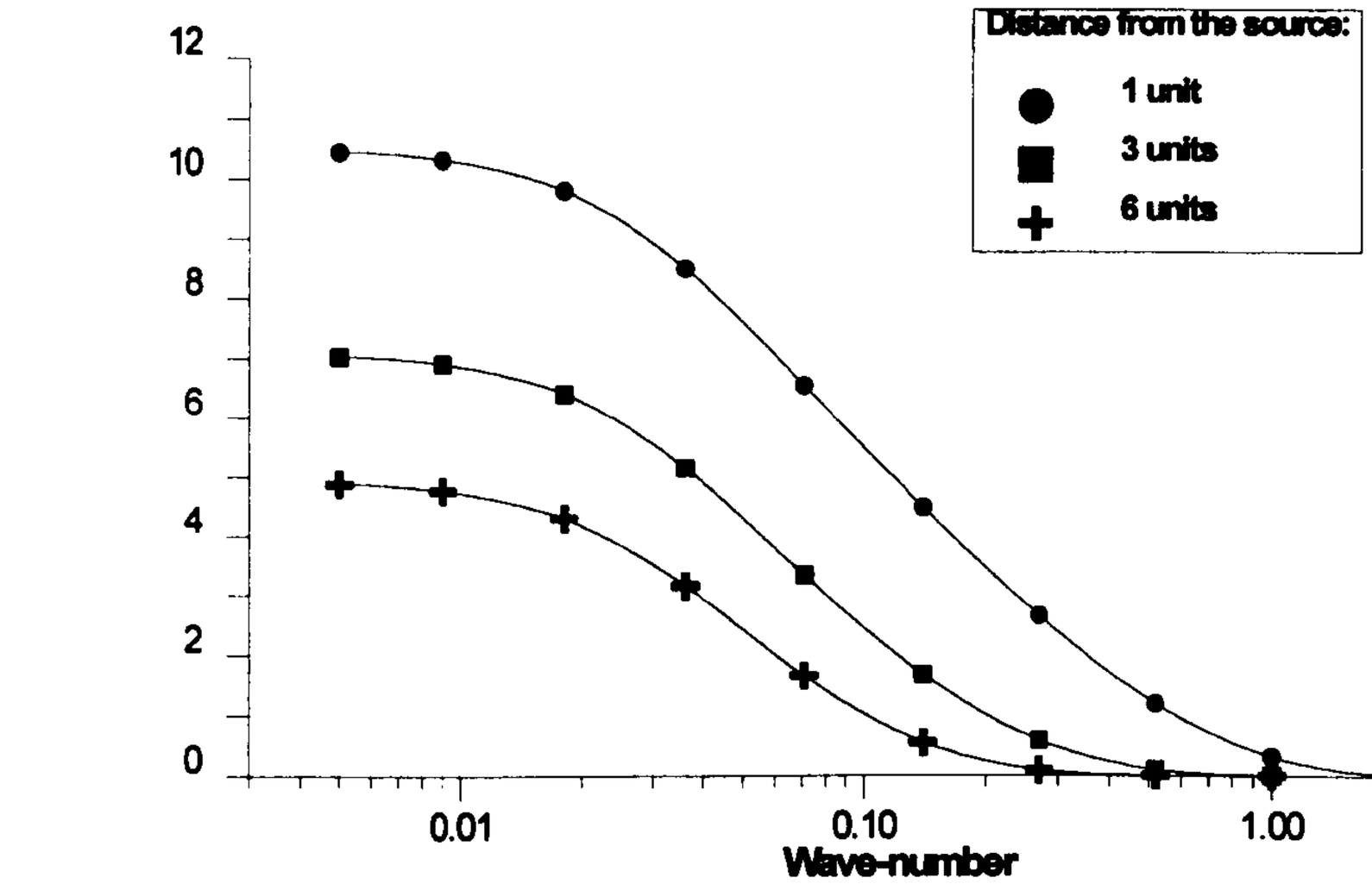


Figure 3.8: The variation of the transformed potential for different distances from the s ource.

the other hand, the modelling scheme is more time-consuming. Studies by Dey and Morisson (1979a) indicated that 6-8 k points are a good compromise between accuracy and time. In the scheme presented here subsectional exponential fitting using 7 wavenumber points was used.

The wave-number values were calculated as follows: starting with an initial (very small) wavenumber value (k_{min}) the choice of the appropriate ending wavenumber value was based on analytical calculations of the transformed potential using equation 3.40. Since the distances (r) between the receiver and the source are known in advance the maximum wavenumber k_{max} was chosen to be such that $\tilde{V}(k_{max}, r) \rightarrow 0$. The remaining wavenumbers were placed at constant intervals in a logarithmic scale in-between the first and last wavenumber values.

When two-electrode arrays (pole-pole) are modelled the singularity of the transformed potential at $k\ y = 0$ introduces errors that are not cancelled out and thus more sophisticated integration techniques have to be considered. For these types of arrays, an alternative approach (in order to avoid the extra computational cost

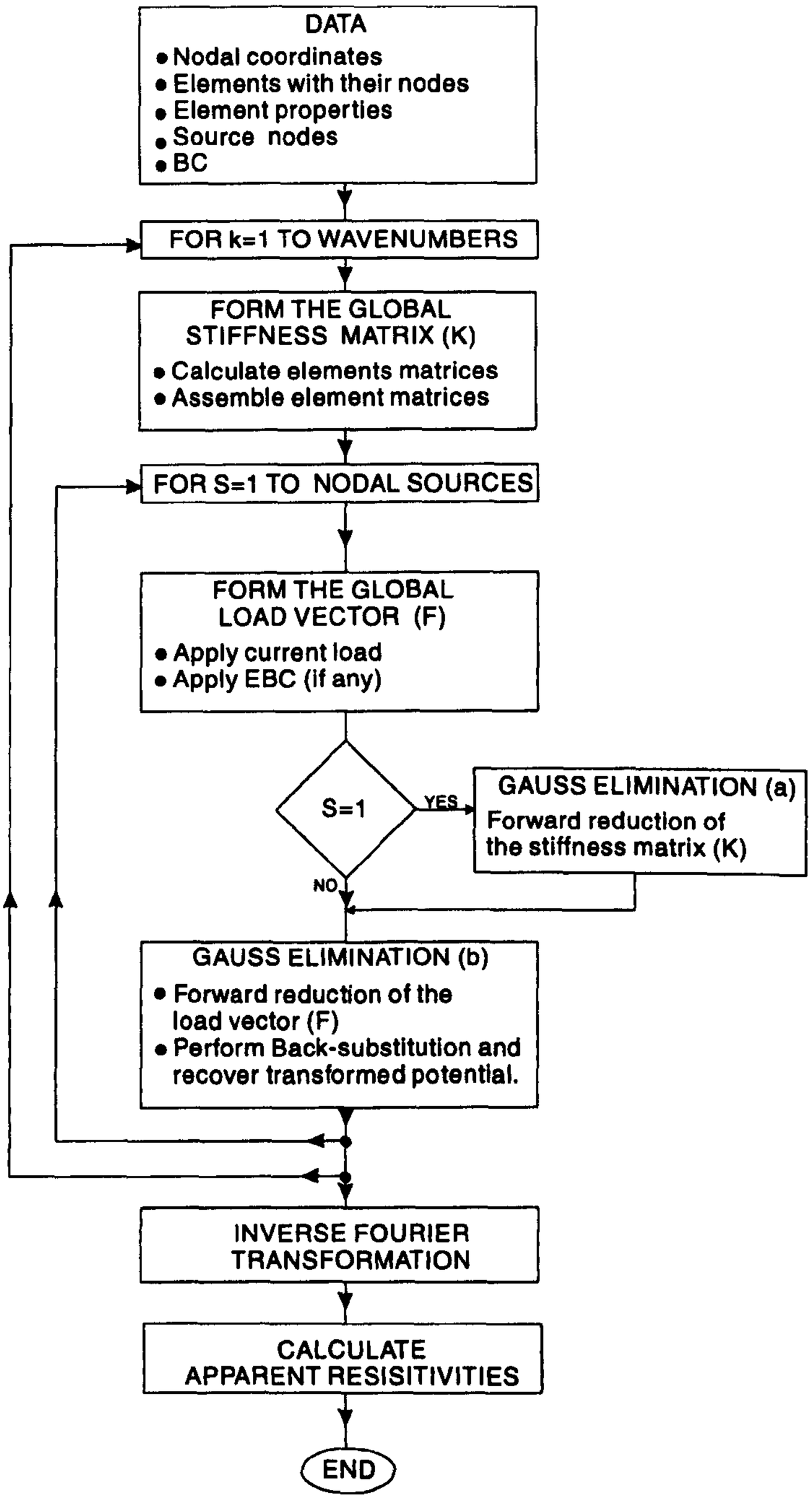


Figure 3.9: A simplified flow-chart of the 2.5D FEM algorithm.

implicit in using more k values) can be followed (Zhao et al., 1986). The response to a homogenous ground can be calculated using the FEM and any anomalous response can then be normalized with respect to the homogenous ground case. In this way increased accuracy can be obtained.

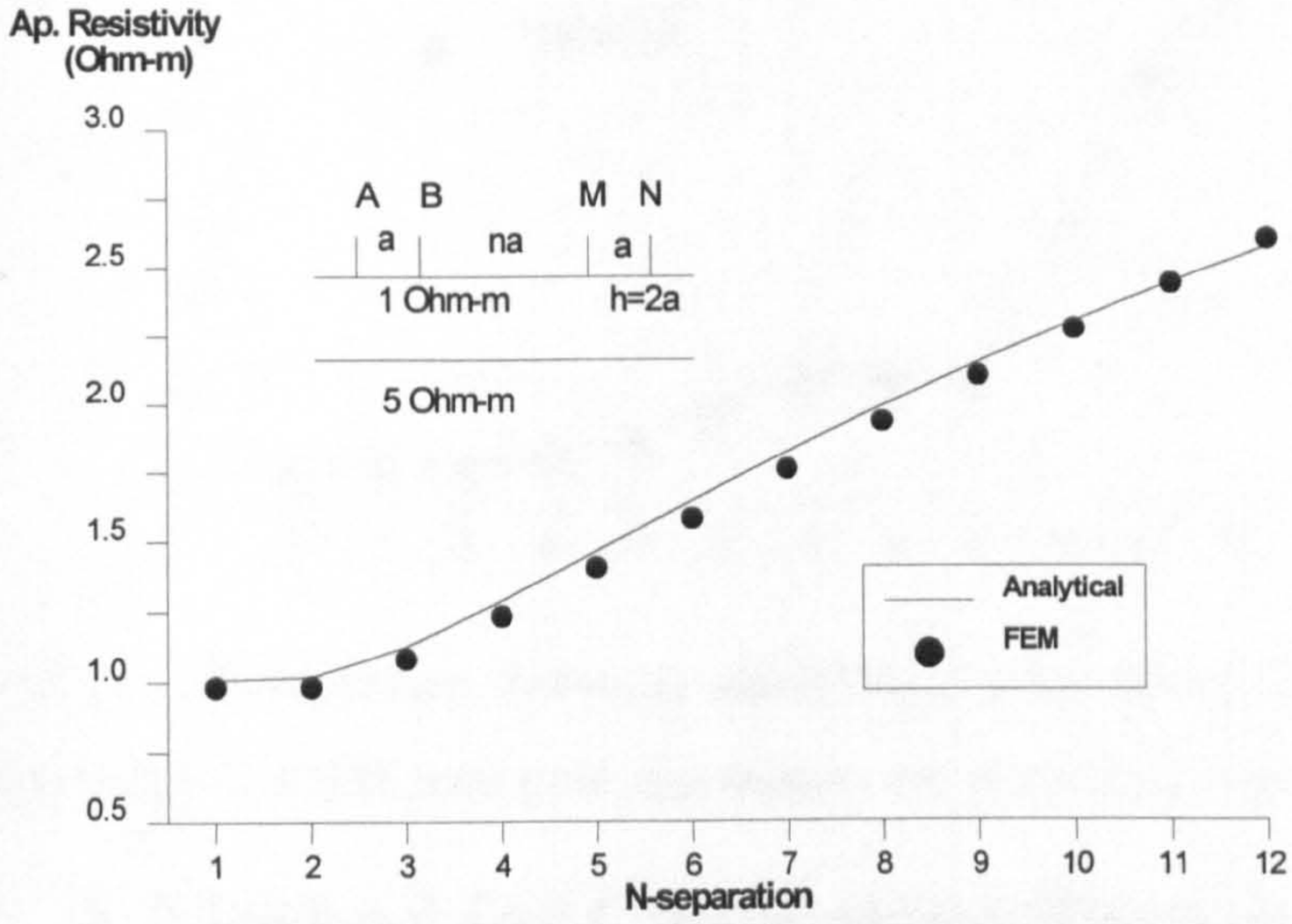
Determination of the Apparent Resistivity - Overall, this scheme, given a resistivity distribution and a current source configuration, will provide the potential values at discrete points, and since a known current (I) is applied the corresponding resistances can be found. In order to calculate the resistance R_{MN} which corresponds to a particular array it is preferable to work with one source each time and to superimpose the effect of these sources (see section 2.1.6).

$$R_{MN} = R_{AM} - R_{AN} - R_{BM} + R_{BN} \tag{3.42}$$

That happens because the variation of the transformed potential due to multiple sources is complicated (Queralt et al., 1991) and therefore more sophisticated interpolation schemes would need to be used.

As long as the resistances have been established the apparent resistivities can be calculated easily using the known formula (see section 2.1.6) $\rho_a = 2\ \pi\ R_{MN}/G$, where G is the geometrical factor of the array.

In Figure 3.9 a simplified flow-chart diagram of the 2.5-D FEM algorithm is depicted. Several comparative tests were carried out in order to verify the validity of the modelling scheme. In Figure 3.10 a comparison is given between analytical and numerical dipole-dipole responses (calculated with the present scheme) from two different models (a 2-layered earth and a vertical discontinuity, also shown in the figure). It can be seen that the agreement is quite good. In Figure 3.11 the normalized and un-normalized pole-pole responses over a vertical discontinuity are shown together with the analytical response. The normalization procedure significantly increased the accuracy of the results.



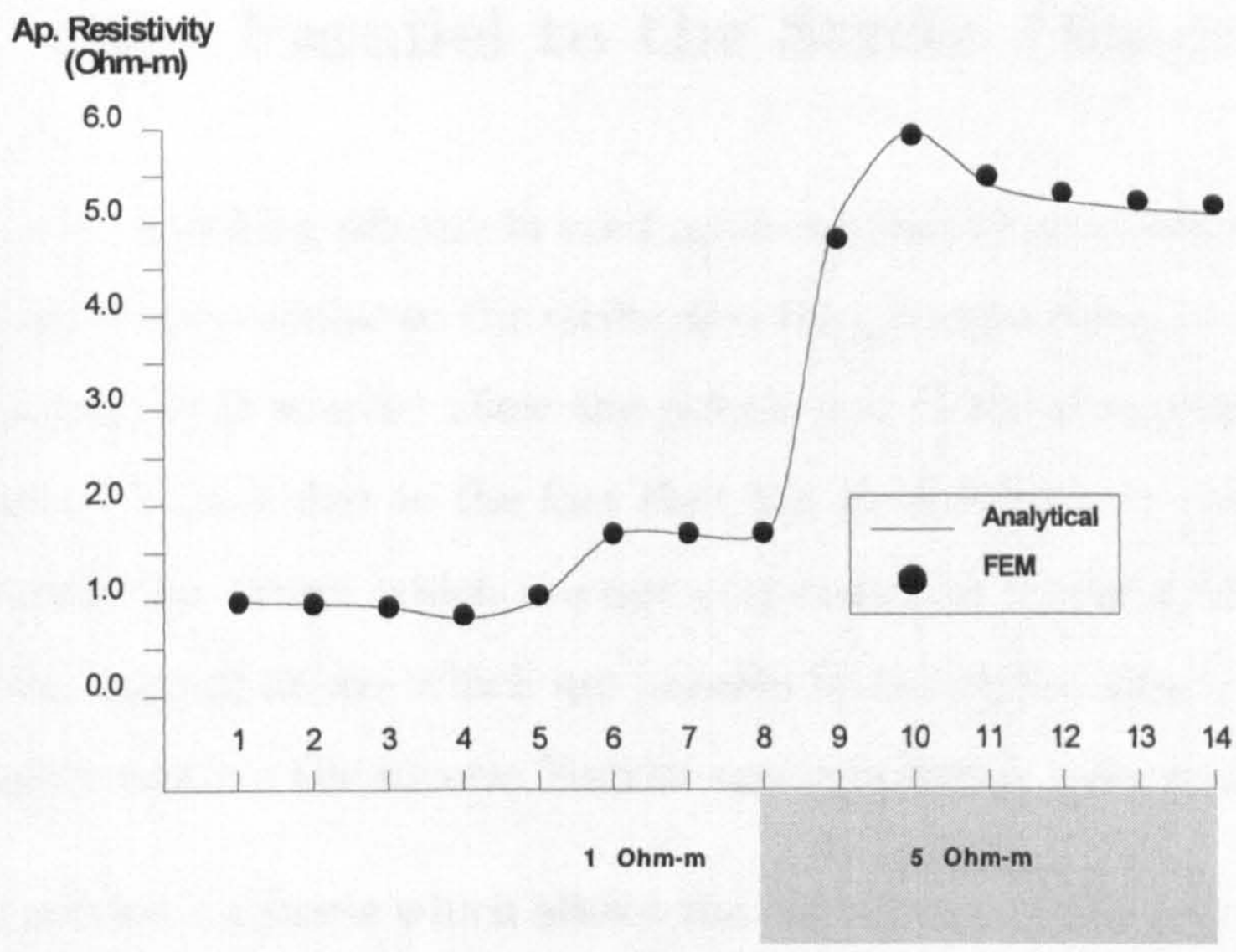


Figure 3.10: Analytical and 2.5-D FEM dipole-dipole responses over a 2-layered earth (top) and an vertical discontinuity (bottom).

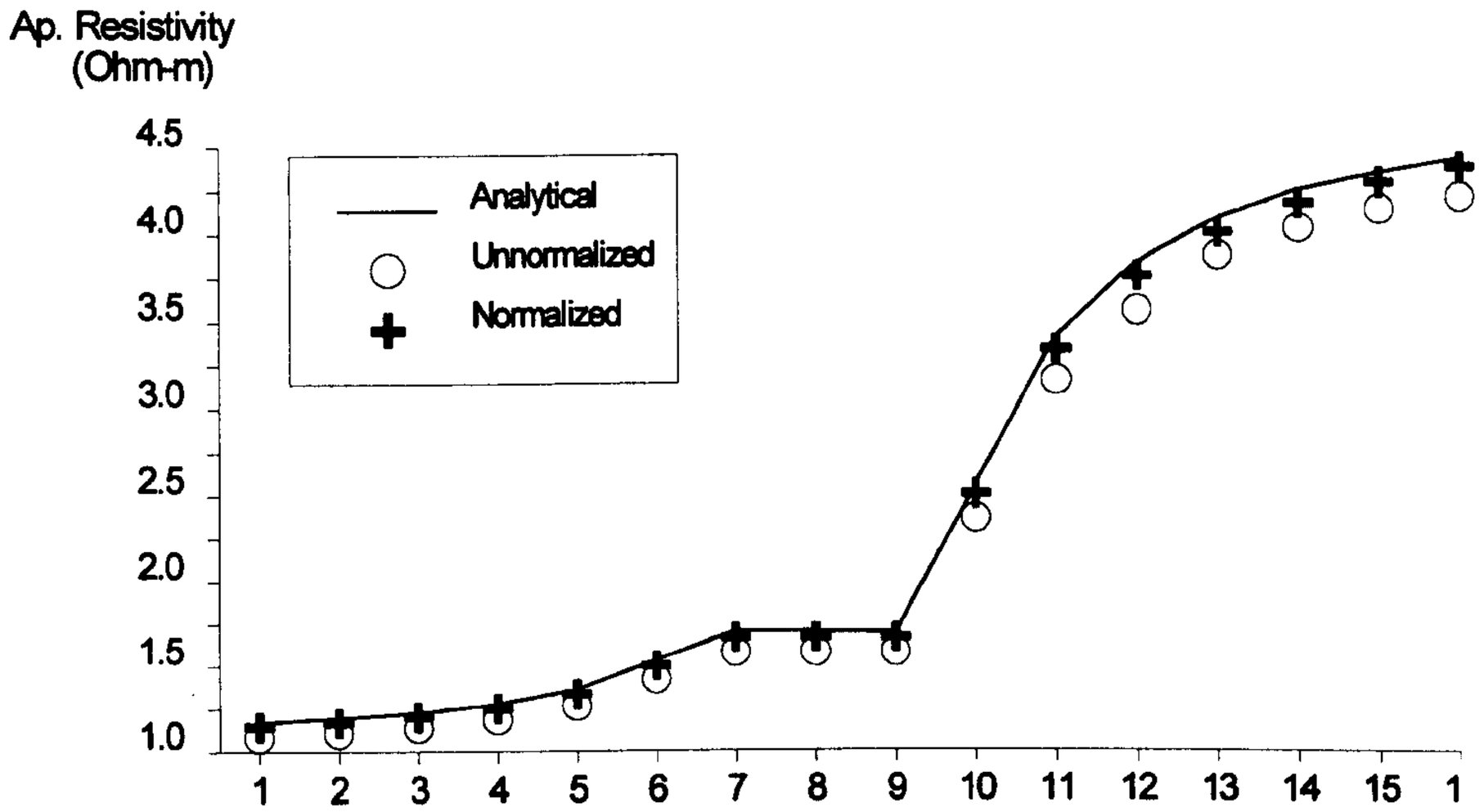


Figure 3.11: Comparison between analytical, normalized 2.5-D FEM and unnormalized 2.5-D FEM pole-pole responses over a vertical discontinuity.

3.4 A Method for Calculating Responses from Arrays Parallel to the Strike Direction

The 2.5-D modelling scheme is used most commonly to provide responses of arrays which are perpendicular to the strike direction, despite the fact that 2.5-D modelling assumptions (3-D source) allow the calculation of the potential at any point in the 3-D space. This is due to the fact that the inverse Fourier transformation is quite complicated for arrays which are not perpendicular to the strike direction. For the particular case of arrays which are parallel to the strike direction the existence of a singularity renders the inverse Fourier transformation even more difficult.

In this section a scheme which allows the calculation of the potential due to a source at any point of a 3-D space is presented. The scheme is a modified version of an algorithm proposed by Queralt et al., 1991. In their work an accurate interpolation scheme is used to perform the inverse Fourier Transform of the potential variations parallel to the strike direction: the total potential is calculated using the FEM, and in order to avoid the singularity problems at the current source, the potential is evaluated at a point very close to the source, with an implicit assumption that the

error is negligible.

This scheme has the disadvantage that it requires extra refinement of the FEM mesh near the current source and therefore is impractical when resistivity profiling is to be considered, since many current positions are involved in a complete profile calculation. In order to avoid this problem, the approach has been modified to avoid the extra mesh refinement without a loss in accuracy. With the presented scheme two things are achieved: **a)** The advantages of the 2.5-D modelling scheme are being exploited in full. **b)** responses of arrays (such as the square array), which previously could not be calculated using the 2.5-D modelling, can now easily be obtained.

The inverse Fourier transformation scheme presented in this section is associated with an alternative strategy for solving the forward problem. This strategy is described in the following subsection.

3.4.1 Solving for the secondary potential

In order to avoid singularity problems related to the existence of the current source it is possible to solve for the secondary potential by assuming that the total transformed potential $\tilde{\mathbf{V}}^T$ is the sum of a primary transformed potential $\tilde{\mathbf{V}}^P$, which derives from a particularly simple model, and a secondary transformed potential $\tilde{\mathbf{V}}^S$ which originates from the superposition of a modelled body which produces deviations from the primary model. Such an approach has been proposed by Rijo (1977) and Lowry et al. (1991).

Here, it is considered that the primary potential is due to a homogeneous half-space and for this case the FEM global system of equations will be (see equation 3.38):

$$\mathbf{K}^P \, \mathbf{A}^P = \mathbf{F} \tag{3.43}$$

Since the source terms in equations 3.38 and 3.43 are identical, we can replace the source vector for the total potential with its equivalent for the primary potential, so

that the global system for the total potential can be expressed as:

$$\mathbf{K}^T \mathbf{A}^T = \mathbf{F} = \mathbf{K}^P \mathbf{A}^P \tag{3.44}$$

Replacing the total transformed nodal potential $\mathbf{A}^T$ by the sum of the primary $\mathbf{A}^P$ and secondary $\mathbf{A}^S$ ones yields:

$$\mathbf{K}^T \mathbf{A}^S = -(\mathbf{K}^T - \mathbf{K}^P) \mathbf{A}^P = -\mathbf{K}^D \mathbf{A}^P \tag{3.45}$$

The matrix $\mathbf{K}^D$, which is the difference between the total and the primary stiffness matrices, has non-zero terms only where the model departs from homogeneity. This matrix can be found readily since it is a function of the properties of each element and the nodal coordinates, while the vector $\mathbf{A}^P$ can be calculated analytically using equation 3.40. The resulting global system of equations (equation 3.45) can be solved in exactly the same manner as the one formed for the total potential. Finally the secondary potential can be recovered by applying the inverse cosine Fourier transform and using numerical integration.

The main advantage of this procedure is that, in effect, the current source is being replaced by fictitious sources placed inside and at the borders of the perturbing body (Rijo, 1977), and in this way the extra refinement of the mesh near to the current source can be avoided. However, the integration of the transformed secondary potential is quite complicated since it usually contains a local maximum (Pelton et al., 1978) but is only sparsely sampled. For arrays which are not perpendicular to the strike direction the interpolation becomes even more difficult because of a cosine term in the inverse Fourier transform. Hence, in our scheme the interpolation is carried out on the total transformed potential by adding the analytically calculated primary transformed potential to the secondary transformed potential.

$$\tilde{V}^T = \tilde{V}^S + \tilde{V}^P = \tilde{V}^S + \frac{I}{2\pi\sigma^P} B_0(k\,r) \tag{3.46}$$

where σ^P is the conductivity of the primary model and $B_0(k\,r)$ is the modified Bessel function of zero order, corresponding to the analytical solution of the transformed

potential for a homogeneous medium: k is the wavenumber and r is the radial distance from the source (independent to the strike direction y, $r = \sqrt{x^2 + z^2}$).

3.4.2 Arrays parallel to the strike direction.

Depending on the orientation of the array relative to the strike direction one extreme case exists as far as the inverse cosine Fourier transform is concerned: the arrays is parallel to the strike direction. In this situation the cosine term remains in equation 3.41, introducing an associated oscillatory effect, and therefore an accurate integration scheme is needed.

The integration method, which is fully described in Queralt et al. (1991), uses a combination of a logarithmic and exponential interpolation for the low and high values of kr, using 10 wavenumber values. This approach was also used in this work, since the total potential is also used as the input to the integration algorithm.

In order to avoid the singularity problems which arise when considering variations parallel to the strike direction but which go through the current source Queralt et al. (1991) calculated the potential at a "small" distance e from the current source, (e should be at least 20 times smaller than the y distance in order for the error to be acceptable) assuming that

$$V_k(x_s, k, z_s) \simeq V_k(x_s + e, k, z_s) \tag{3.47}$$

where x_s, z_s are the coordinates of the source. As far as the FEM is concerned, the existence of a node so close to the current source implies that a local mesh refinement should be performed - rendering the scheme impractical when resistivity profiling schemes are to be modelled. Moreover, no matter the detailed form of the refinement, the existence of tiny elements next to the source can increase the ill-conditioning of the FEM stiffness matrix (Burnett, 1989).

In the modified scheme used here, the singularities arise as well in the sense that the modified Bessel function $B_0(k\ r)$ has no solution when r=0, that is for points which

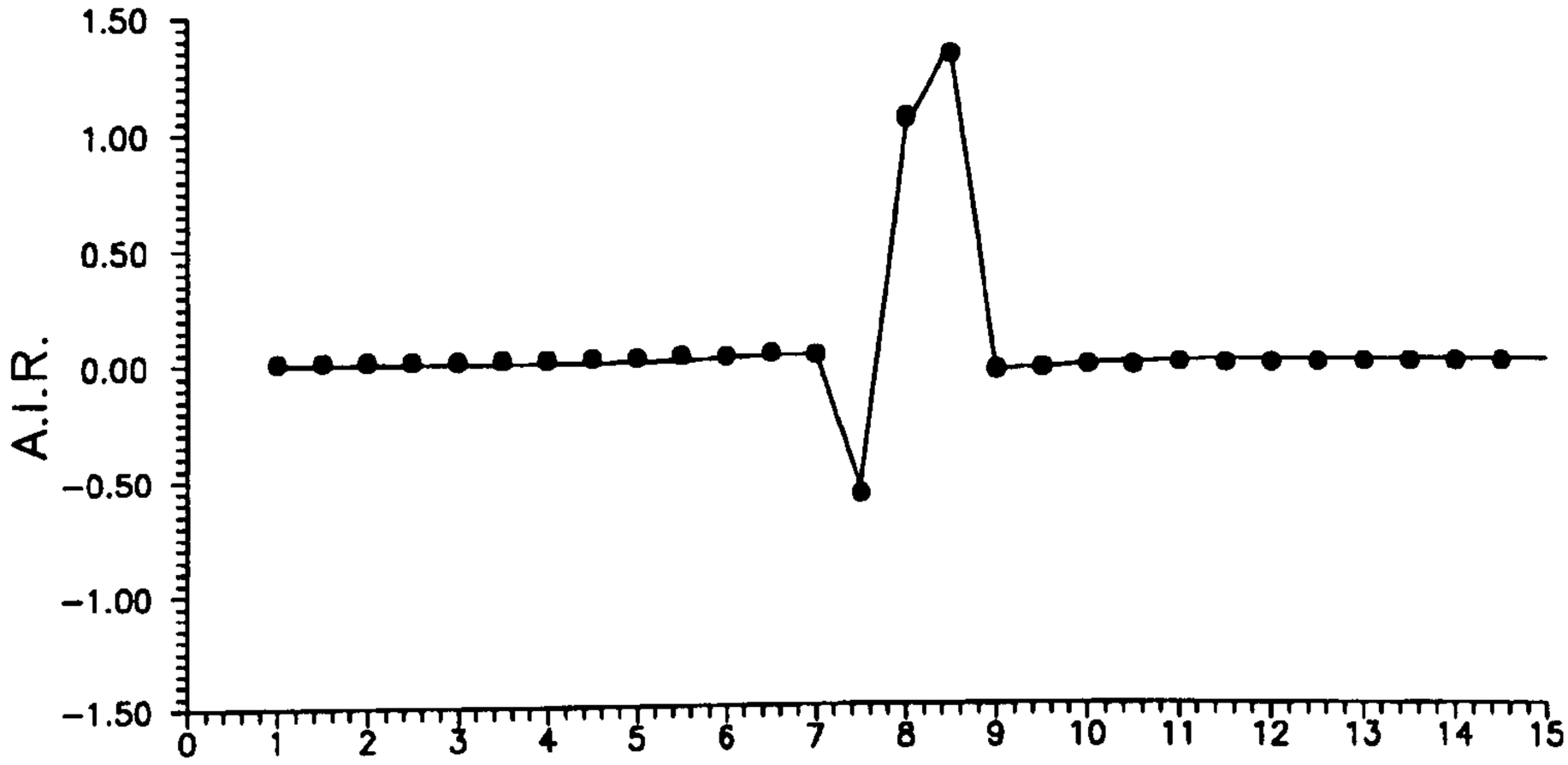


Figure 3.12: Analytical (continuous line) versus numerical (circles) response of the AIR over a vertical discontinuity measured every half unit.

are in line with the current source in a direction parallel to the strike. Making a similar assumption to that of Queralt et al. it is assumed that:

$$\begin{aligned} \mathbf{A}^T(x_s, k, z_s) &\simeq \mathbf{A}^S(x_s, k, z_s) + \mathbf{A}^P(x_s + e, k, z_s) \\ &= \mathbf{A}^S(x_s, k, z_s) + \frac{I}{2\pi\sigma^P} B_0(k\ e) \end{aligned} \tag{3.48}$$

In this way, without a loss in accuracy, it is possible to avoid the extra mesh refinement. For any other orientation of the array relative to the strike direction the inverse cosine Fourier transform can be calculated using the same scheme without this assumption.

Several comparative tests were carried out in order to verify the validity of the modelling scheme. In Figure 3.12, a comparison between the analytical and numerical square array AIR response over a vertical discontinuity is presented, with good agreement. The square array is considered to have a side of one unit while the array step is half a unit.

3.4.3 Response of common 2-D structures

The effect produced by various buried structures was calculated using the scheme outlined above. The model structures were designed to be representative of features that occur commonly in field exploration. The measurement profile was considered to be perpendicular to the strike.

The scheme allowed the calculation of the square array α, β configurations and AIR responses. Responses from others arrays were also calculated and are presented here. The sizes of the side of the square and the Wenner array spacing were set to one (arbitrary) unit. The spacing of the pole-pole probes and the potential dipole for the Schlumberger gradient, and dipole-dipole arrays were also set to unity.

Figure 3.13a shows the calculated effect when profiling over a vertical contact, as illustrated. The two media have a resistivity contrast of ten. The AIR response shows a clear dipolar signal which contrasts with the zero response on both sides of the contact.

A model which is particularly appropriate for archaeological geophysics is depicted in Figure 3.13b together with the calculated anomalies: a vertical square prism of side-dimension two units is buried at a depth of one sampling interval in a medium which is ten times more conductive. The body is assumed to simulate the remains of a buried wall or a tomb. On a geological scale, it could represent an uplift of the bedrock, an anticline or even an intrusion. The AIR anomaly has a distinctive pattern which can be seen to be similar to that of a synthesis of two vertical contact responses of opposite sign. The edges of the disturbing body can be defined to a reasonable approximation by the turning points of the signal. The Schlumberger gradient, the dipole-dipole (Wenner β) and the square β arrays give fairly restricted spatial signals. On the contrary, the square α, Wenner α and pole-pole arrays give relatively smooth and wide anomalies.

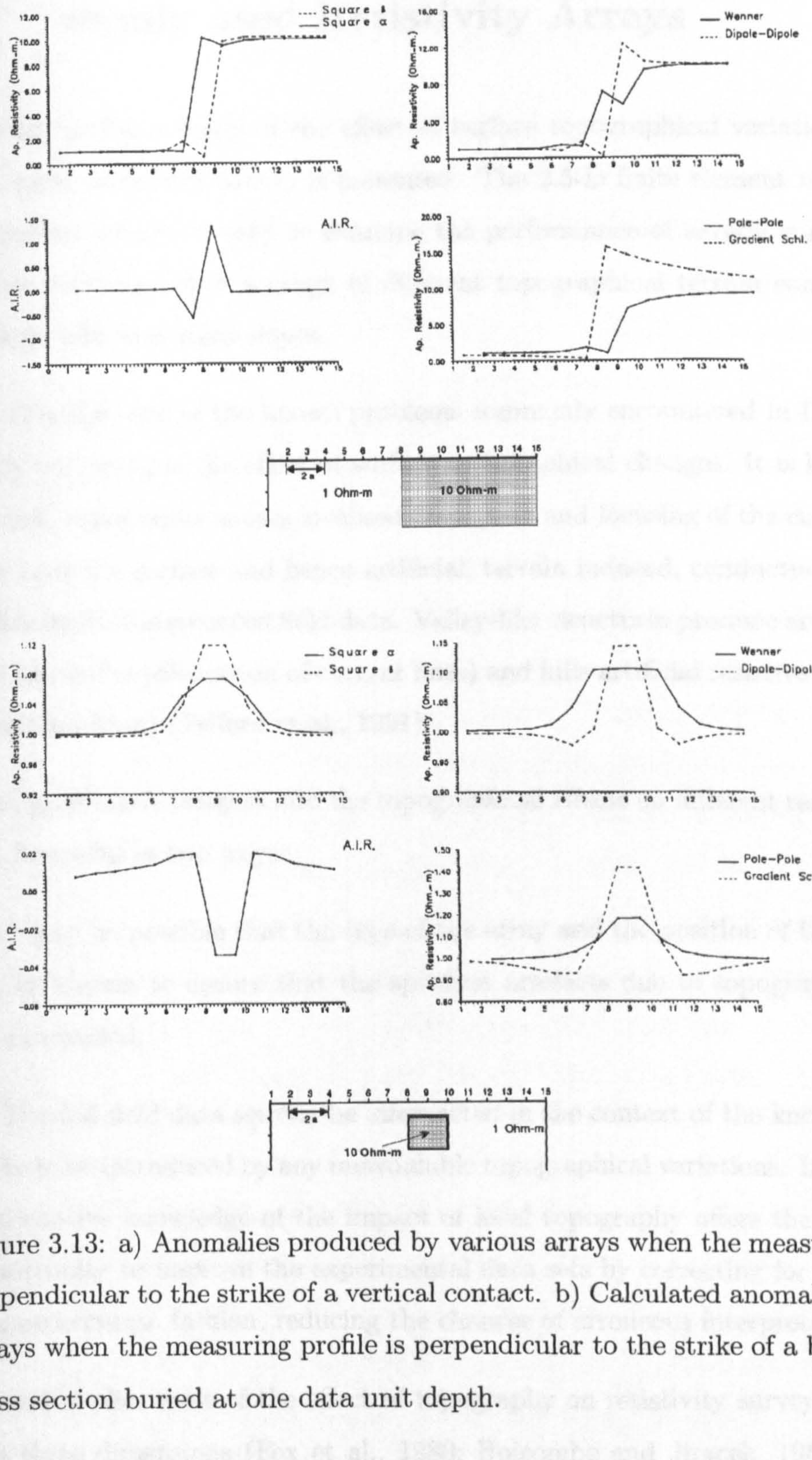


Figure 3.13: a) Anomalies produced by various arrays when the measuring profile is perpendicular to the strike of a vertical contact. b) Calculated anomalies for various arrays when the measuring profile is perpendicular to the strike of a body of square cross section buried at one data unit depth.

3.5 The Effect of Terrain Topography on Commonly used Resistivity Arrays

In this section a study of the effect of surface topographical variations on several DC earth resistivity arrays is presented. The 2.5-D finite element method (FEM) modelling scheme is used to examine the performance of several survey arrays on buried features below a range of different topographical terrain contexts, such as valleys, hills and steep slopes.

At all scales, one of the known problems commonly encountered in DC earth resistivity surveying is the effect of surface topographical changes. It is known that, in general, topography causes localised dispersion and focusing of the current distribution near the surface and hence artificial, terrain induced, conductive and resistive anomalies in the recorded field data. Valley-like structures produce artificial conductive anomalies (dispersion of current lines) and hills artificial resistive ones (focusing of current lines) (Telford et al., 1991).

Even qualitative insights into the topographical effects on different resistivity arrays can be useful in two ways:

a) It may be possible that the type of the array and the position of the survey lines can be chosen to ensure that the spurious artefacts due to topographical changes are minimized.

b) The full field data set can be interpreted in the context of the known distortions likely to be introduced by any unavoidable topographical variations. In fact, detailed quantitative knowledge of the impact of local topography offers the more enticing opportunity to improve the experimental data sets by correcting for topography in a more accurate fashion, reducing the chances of erroneous interpretations.

Several studies exist of the effect of topography on resistivity surveys, both in two and three dimensions (Fox et al., 1980; Holcombe and Jiracek, 1984) but are all

limited to the case of the dipole-dipole array. Methods for the quantitative correction of resistivity data sets which contain "noise" due to topographical variations have also been proposed (Fox et al., 1980; Tong and Yang, 1990)

Unfortunately, however, studies of other array types and any comparative studies of the performance of different arrays in varying field contexts do not exist in the published literature (Ward 1989). In this work, the FEM forward modelling scheme is used to test several widely-used resistivity arrays, namely dipole-dipole, Wenner, pole-pole, pole-dipole and the Square array. This short study involves a series of calculations of both profiling and pseudosection surveys, with different probe configurations, each in the context of a range of commonly encountered idealized topographical models: valleys, hills and steep slopes.

3.5.1 Incorporation of the topography into the FEM scheme

In order to model the topographical effects a FEM mesh which is consistent with the terrain variations has to be used. An easy way to achieve this is to generate a rectangular mesh with x coordinates which are consistent with the electrode positioning and then just raise or lower the depth (z-coordinates) of the surface nodes to match the terrain. However, abrupt terrain changes will cause the elements related to the surface nodes to be very large in the z direction - this can dramatically affects the accuracy of the scheme, since discretization near the sources should be high. An example of the mesh produced using this technique for a simple terrain model can be seen in Figure 3.14b.

An automatic mesh-generator was developed to produce a FEM mesh which is consistent with both the lateral and vertical positioning of the probes, in view of the needs (boundary shape) of the particular topography to be studied and without reducing the nodal density at the top of the FEM mesh. This is achieved by creating a rectangular mesh which has basic x coordinates consistent with the x coordinates of the probes and then by adding as many nodal rows in the depth (z) direction as

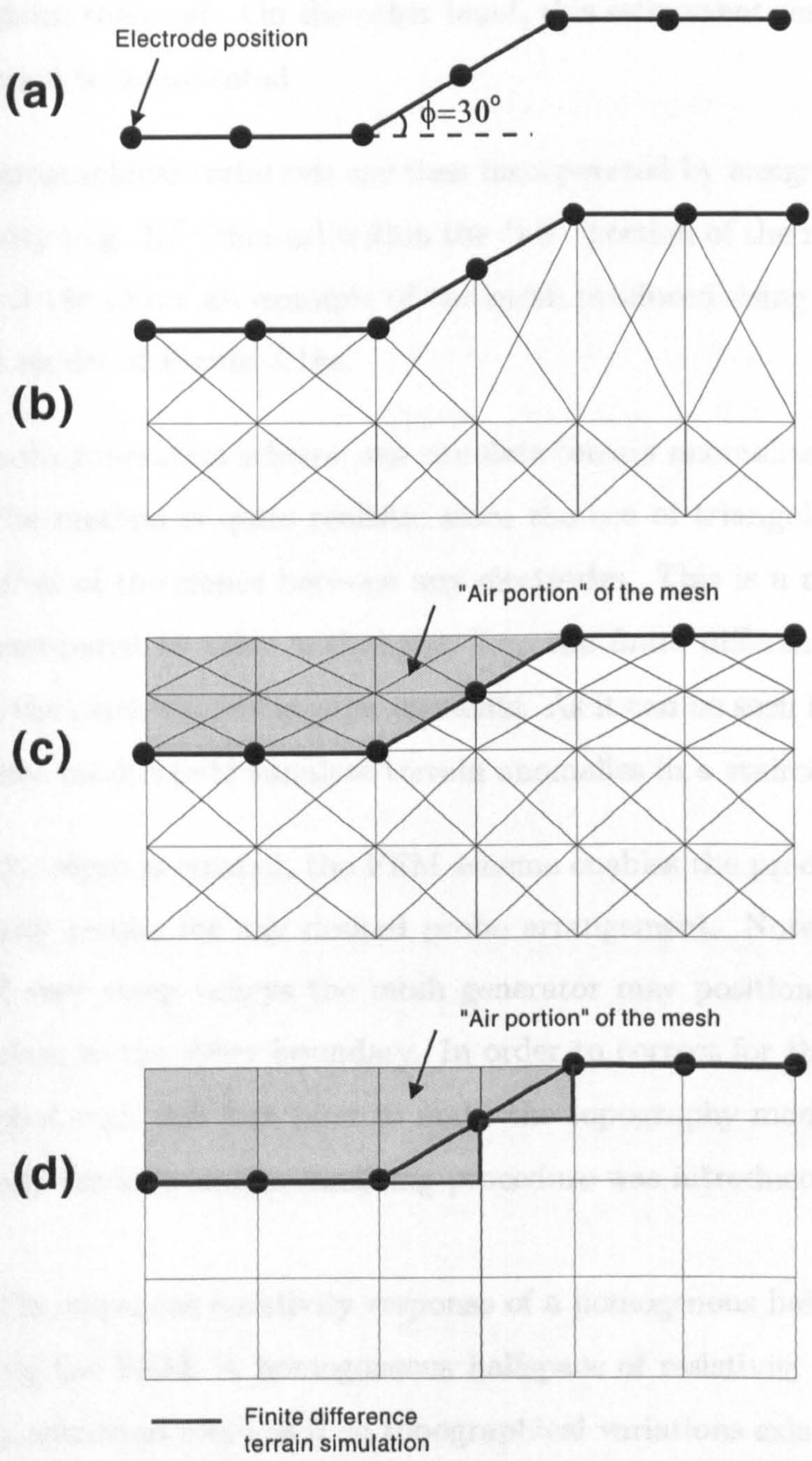


Figure 3.14: Different meshes which incorporate topography: a) the terrain variation to be simulated, b) FEM mesh applying the terrain variations directly to the nodes at the top b) FEM mesh produced by a more sophisticated mesh generator, c) equivalent Finite difference mesh.

there are different probe heights. In this way high a nodal refinement is achieved throughout the mesh. On the other hand, this refinement results large mesh sizes if the terrain is complicated.

The topographical variations are then incorporated by assigning an extremely large resistivity (e.g. 10^6 Ohm-m) within the "air" portion of the mesh (Fox et al., 1980). Figure 3.14c shows an example of the mesh produced using that technique for the terrain model of Figure 3.14a.

This mesh generation scheme can simulate terrain anomalies in great detail. Moreover, the method is quite realistic since the use of triangular elements allows the simulation of the slopes between any electrodes. This is a major advantage of the FEM compared to other techniques (i.e. the finite difference method) which discretize the earth via rectangular elements. As it can be seen in Figure 3.14d, a finite difference mesh would simulate terrain anomalies in a staircase fashion.

Once the mesh is created, the FEM scheme enables the production of the apparent resistivity results for any desired probe arrangement. Note however, that for the case of very steep valleys the mesh generator may position the measuring probes quite close to the lower boundary. In order to correct for the possible inaccuracies associated with this fact (and to make the topography modelling scheme valid for all cases) the following normalizing procedure was introduced:

- The apparent resistivity response of a homogenous halfspace is calculated using the FEM. A homogeneous halfspace of resistivity ρ_o is simulated by the constructed mesh as if no topographical variations existed - no high resistivity is assigned to the "air" portion of the mesh. In this way some of the measuring electrodes appear to be "buried". Each calculated apparent resistivity (P_i) is compared with the known homogenous resistivity and a modelling error is obtained. The error for the measurement i is given by: $e_i = \rho_o / P_i$.

- The "air" portion (and any subsurface structure) is introduced into the mesh

and now the "topography contaminated" apparent resistivities Pt_i are calculated once again.

- Finally the normalized "topography contaminated" apparent resistivity Ptn_i is obtained by: $Ptn_i = e_i Pt_i$

This scheme is almost identical to the normalization procedure for the pole-pole array proposed by Zhao et al. (1986) (it was presented in section 3.3.3). This procedure is based on the assumption that the modelling errors are linear and, although this is a rough approximation, it produces a reasonable correction factor, as was demonstrated in section 3.3.3.

3.5.2 Examples

Several terrain models were tested in order to check their effect on the responses obtained from different resistivity arrays. Some of the results are presented and discussed in this section. It has to be noticed that no direct quantitative comparison between the arrays can be made since no equivalent form of the arrays can be established: Arrays tend to have different sensitivities and thus arrays which are comparatively more sensitive to topographical variations will also be more sensitive in picking up the targets.

Vertical slope model

This model was chosen to simulate a common situation encountered in resistivity surveys of archaeological sites: sometimes surveys have to be conducted in the vicinity of excavation pits or steep slopes (such an example is the Europos acropolis site depicted in Figure 2.14). The response of several arrays commonly used for archaeological site surveys is depicted in Figure 3.15. The model used is also shown [Figure 3.15(bottom)].

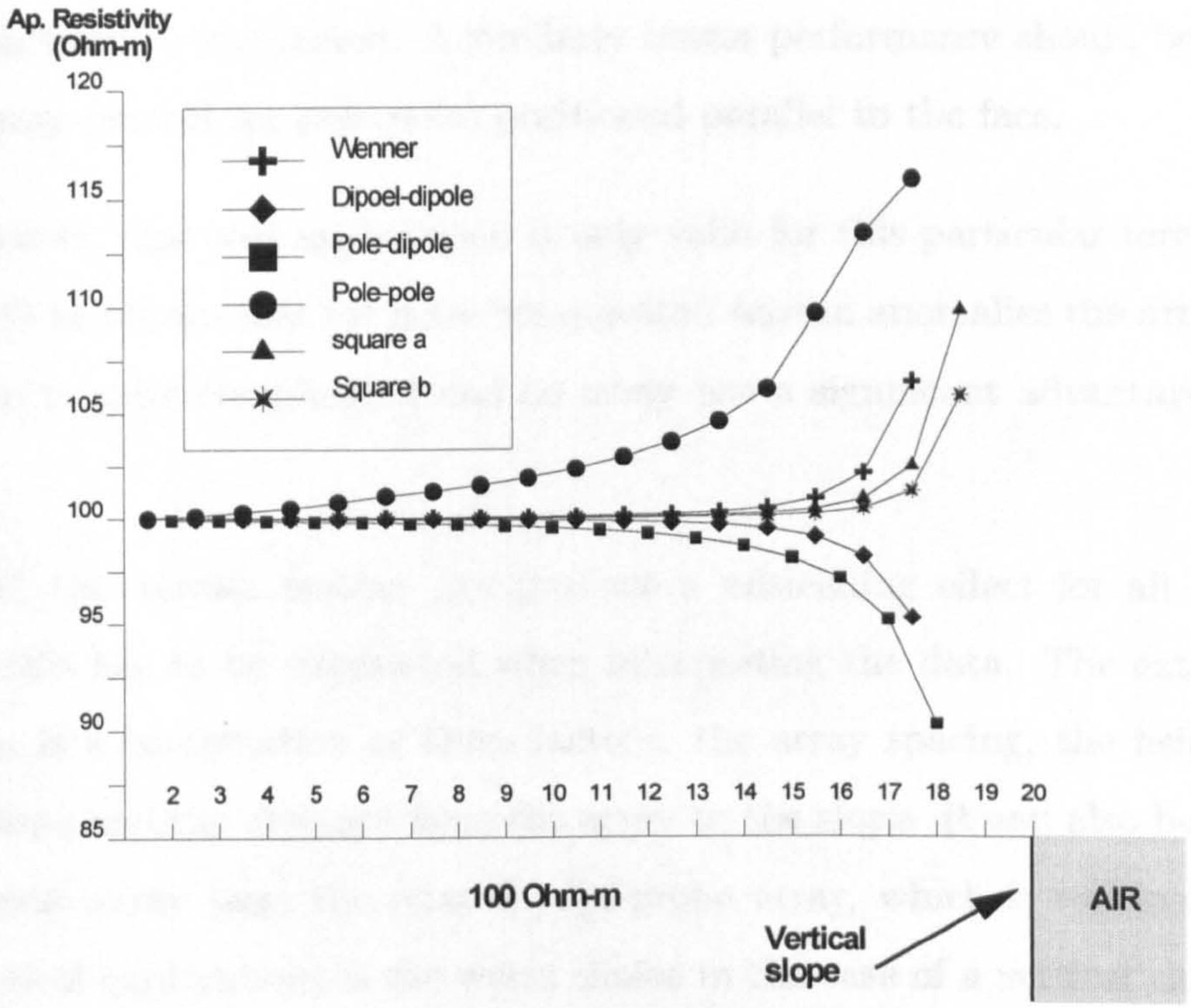


Figure 3.15: Topography effect of a vertical slope (bottom) in the apparent resistivity profiles of different arrays.

The results indicate that the pole-pole array is the most sensitive to this particular terrain type producing an artificial high anomaly. Pole-dipole, Wenner and dipole-dipole arrays are quite sensitive as well. Square α, β arrays are less sensitive.

The results can be explained by considering the configuration of the arrays. The terrain anomaly causes a distortion of the current lines and therefore of the measured potential field. Since pole-pole array has one of the potential electrodes at "infinity" only the one potential probe "senses" the anomaly, and therefore the recorded potential difference is high. Conversely for the remaining arrays both potential electrodes "sense" the anomaly, to some degree, so that the potential difference is affected rather less by the vertical face.

The square array is less sensitive since it is quite compact. In particular the square β array is the least sensitive since both potential electrodes are placed orthogonally to the strike (measuring) direction (parallel to the slope) and hence the measured

effect is partially counteracted. A similarly better performance should be expected for any array (except for pole-pole) positioned parallel to the face.

Note, however, that this explanation is only valid for this particular terrain model since it will be shown that for more complicated terrain anomalies the array-terrain interaction is quite complicated and no array has a significant advantage over the other.

In general, this terrain feature can produce a misleading effect for all of the arrays and this has to be considered when interpreting the data. The extent of the corruption is a combination of three factors: the array spacing, the height of the vertical slope and the distance from the array to the slope. It can also be said that the pole-pole array (and the related twin-probe array, which is routinely used in archaeological exploration) is the worst choice in the case of a vertical slope.

Hill and Valley models

In Figure 3.16 the topographical effect of a hill model (Figure 3.16a) on the response of various arrays is shown for full 2-D data sets (a homogeneous ground is considered). The results are presented in a pseudosection form as grey scale images. In figure 3.17 the % error introduced by this terrain model can be shown. The results indicate that all arrays tend to generate high and low "artificial" anomalies which can be mistakenly attributed to resistive or conductive bodies, or can mask the response of "real" bodies.

The dipole-dipole array (Figure 3.16b) produced a central high apparent resistivity anomaly below the area of the hill which widens as the n separation increases. On either side of the spurious high, further low apparent resistivity regions appear.

The Wenner array (Figure 3.16c) produces an effect opposite to that of the dipole-dipole array: a low apparent resistivity region below the central area of the hill, with a high at either side.

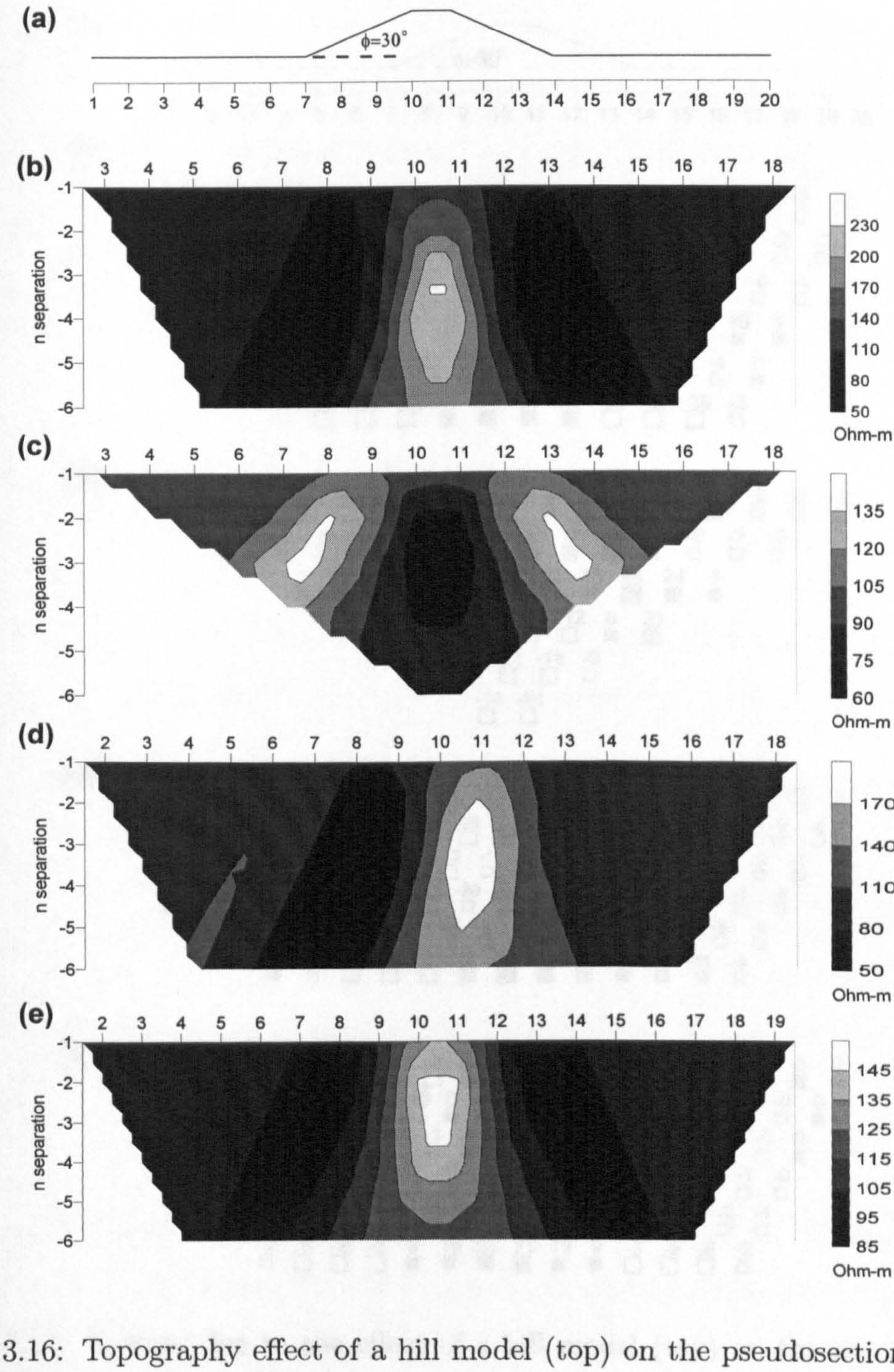


Figure 3.16: Topography effect of a hill model (top) on the pseudosection presentation of a 2-D resistivity survey over a homogeneous ground: b) dipole-dipole, c) Wenner, d) pole-dipole, e) pole-pole.

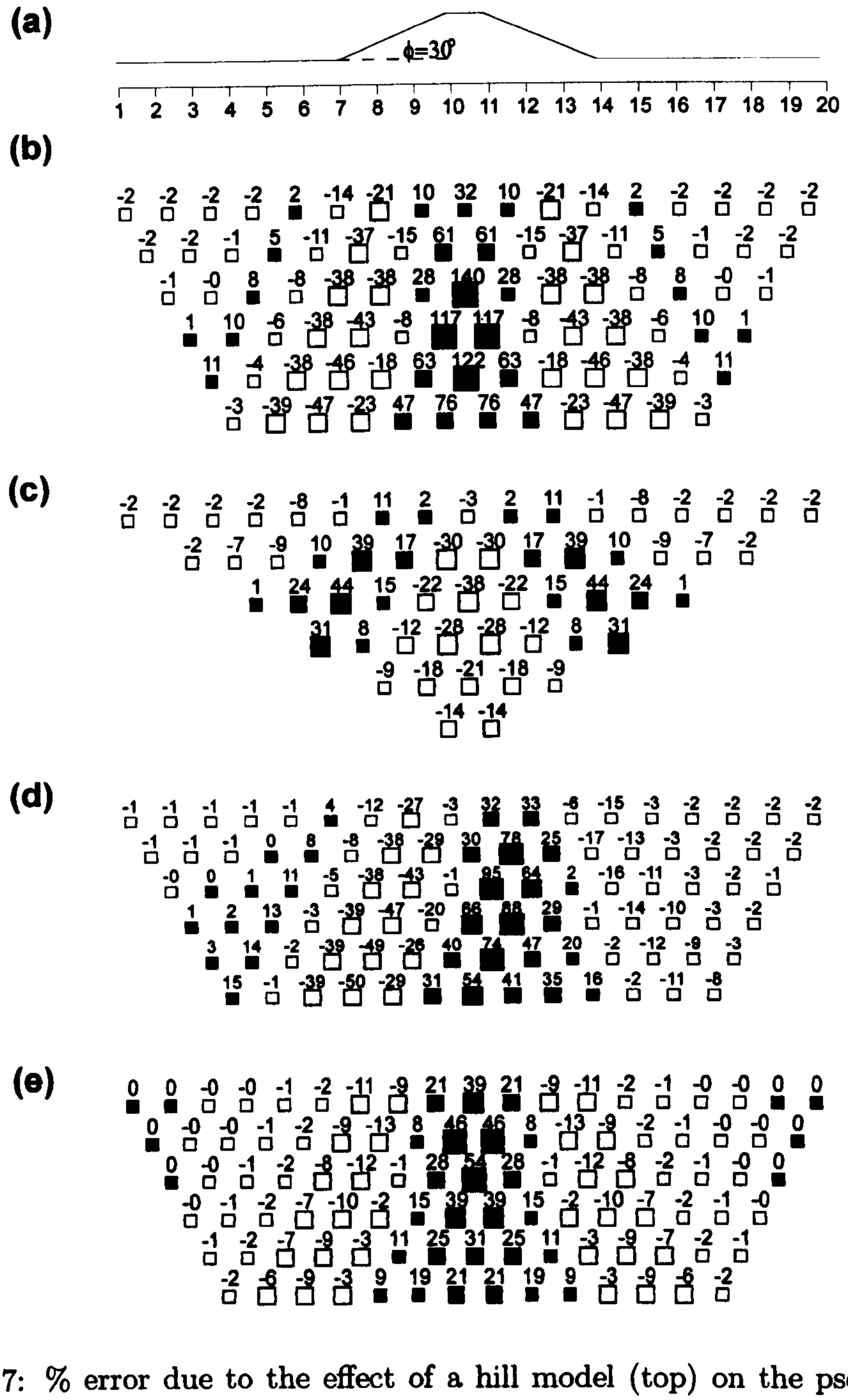


Figure 3.17: % error due to the effect of a hill model (top) on the pseudosection presentatio of a 2-D resistivity survey over a homogeneous ground: b) dipole-dipole, c) Wenner, d) pole-dipole, e) pole-pole.

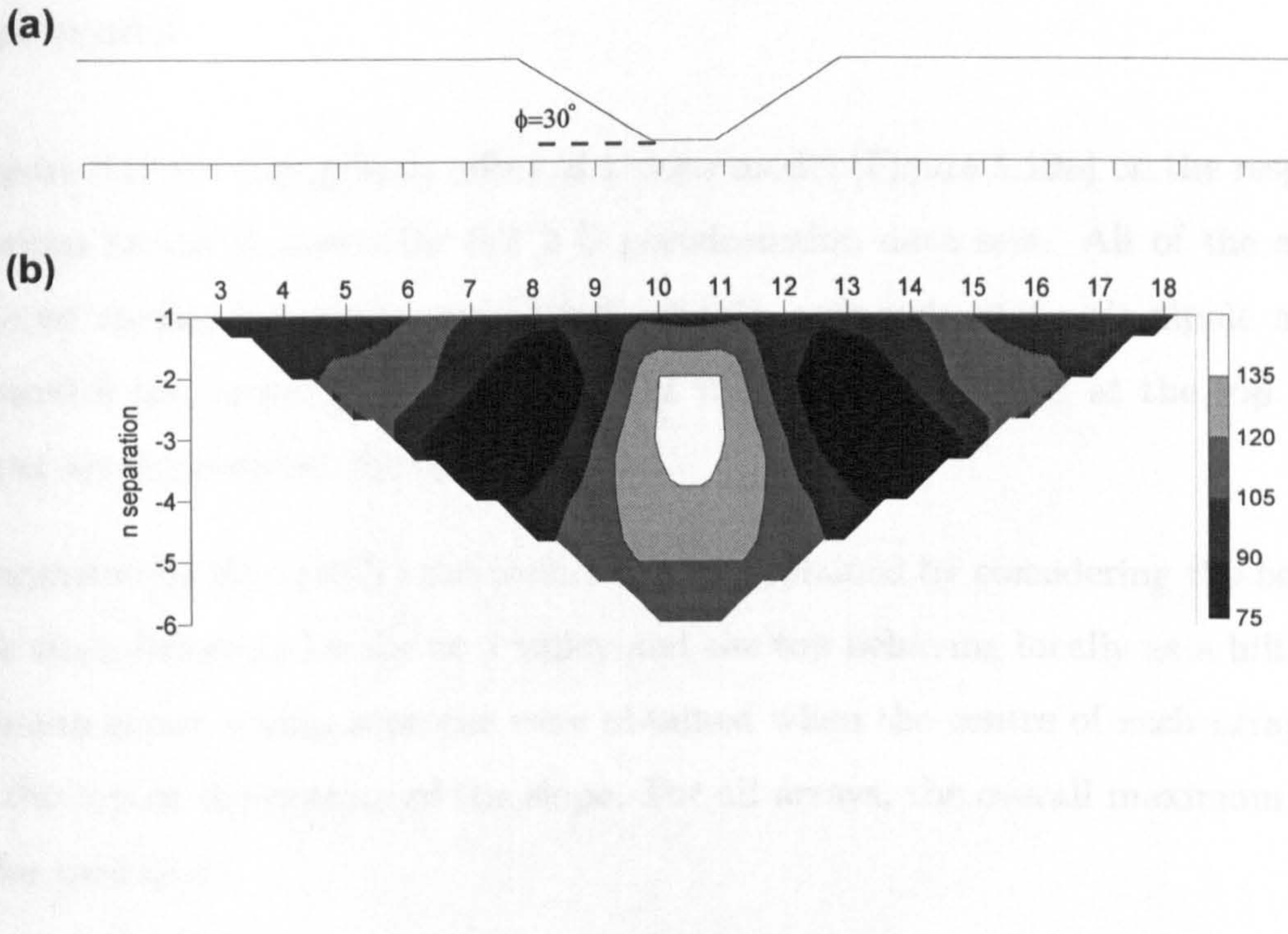


Figure 3.18: Topography effect for a valley model (top) on the pseudosection presentation of a 2-D Wenner data set over a homogeneous ground.

The pole-pole array (Figure 3.16d) produced an anomaly which has a pattern similar to that of the dipole-dipole. The pole-dipole array (Figure 3.16e) produced an asymmetric anomaly with a high region just below and on the right of the hill and a low on the left of the central region [11].

The maximum positive and negative error within a profile was obtained when the centre of the array was positioned at the top of the hill. For all arrays the overall maximum error (both positive and negative) was obtained for probe separation n=3.

As would be expected, the equivalent valley model (Figure 3.18a) produced results directly opposite to those of the hill model. In Figure 3.18b the Wenner array response is depicted.

[11]The probe arrangement of the pole-dipole array is the one shown in Figure 2.8d. If different arrangement was used, i.e. N-M-A instead of A-M-N, the anomaly produced would have had an inverted pattern.

Slope model

In Figure 3.19 the topography effect of a slope model (Figure 3.19a) on the response of various arrays is shown for full 2-D pseudosection data sets. All of the arrays produced asymmetric responses. Dipole-dipole, pole-pole and pole-dipole arrays produced a low response at the bottom of the slope and a high at the top. The Wenner array produced the opposite effect.

As suggested by Rijo (1977) the results can be explained by considering the bottom of the slope behaving locally as a valley and the top behaving locally as a hill. The maximum errors within a profile were obtained when the centre of each array was over the top or the bottom of the slope. For all arrays, the overall maximum error was for n=5 or 6.

3.5.3 Correction of topographical effects

It has been demonstrated that the terrain anomalies can be a significant source of distortion in resistivity surveys. Fox et al. (1980) proposed an approximate way to take the quantitative effect of terrain noise into consideration, and eventually to correct the data.

The technique is quite simple. Assuming that a data set has been obtained over a known field topography, the steps that have to be followed are as follows:

- The topography is simulated and a similar model data set is obtained for a homogeneous ground.
- The % errors introduced within this data set by the purely topographical variations are calculated and stored.
- The original data set is then corrected using the error factors obtained for homogeneous earth.

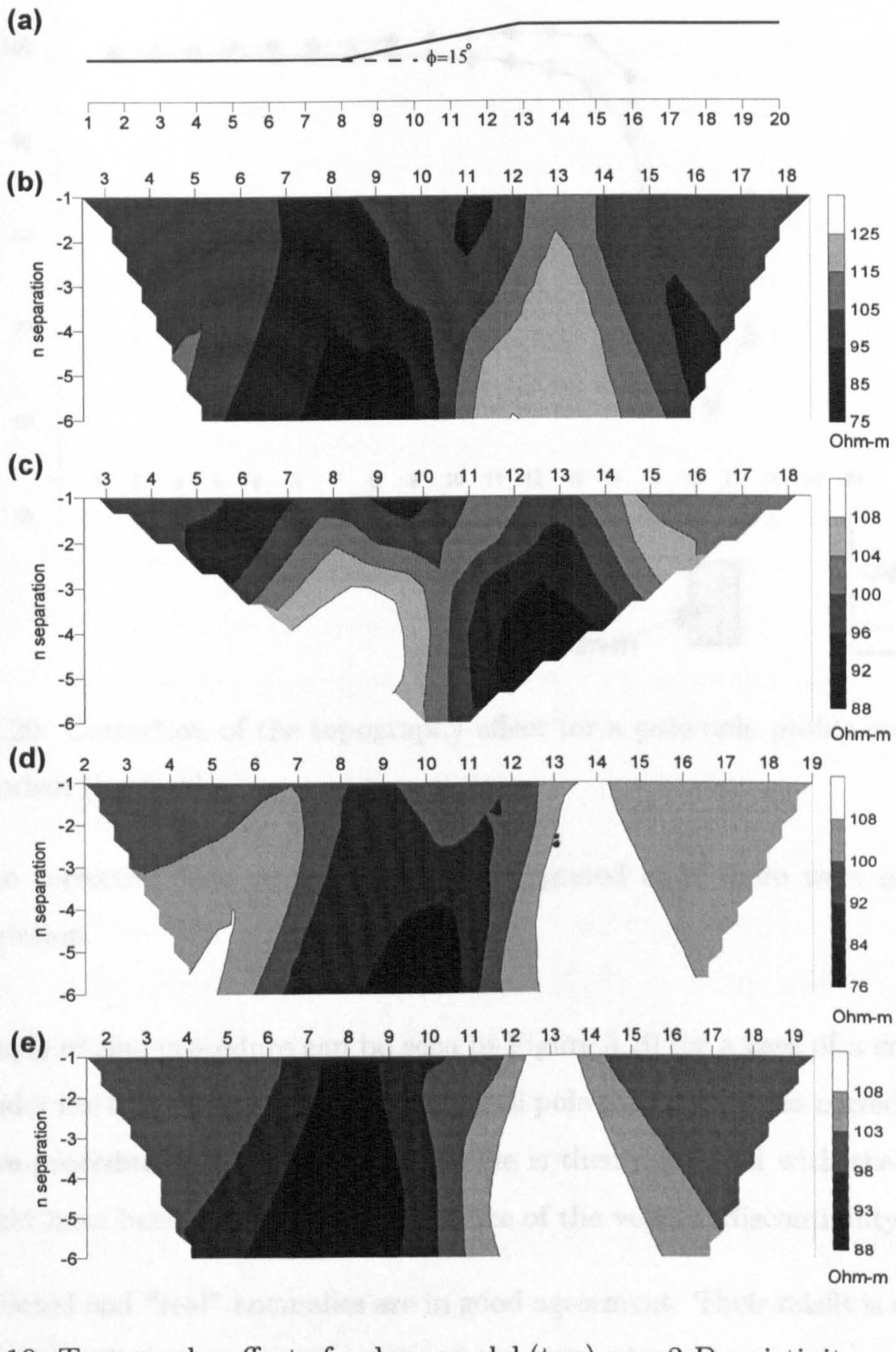


Figure 3.19: Topography effect of a slope model (top) on a 2-D resistivity survey over a homogeneous ground (pseudosection presentation): b) dipole-dipole, c) Wenner, d) pole-dipole, e) pole-pole.

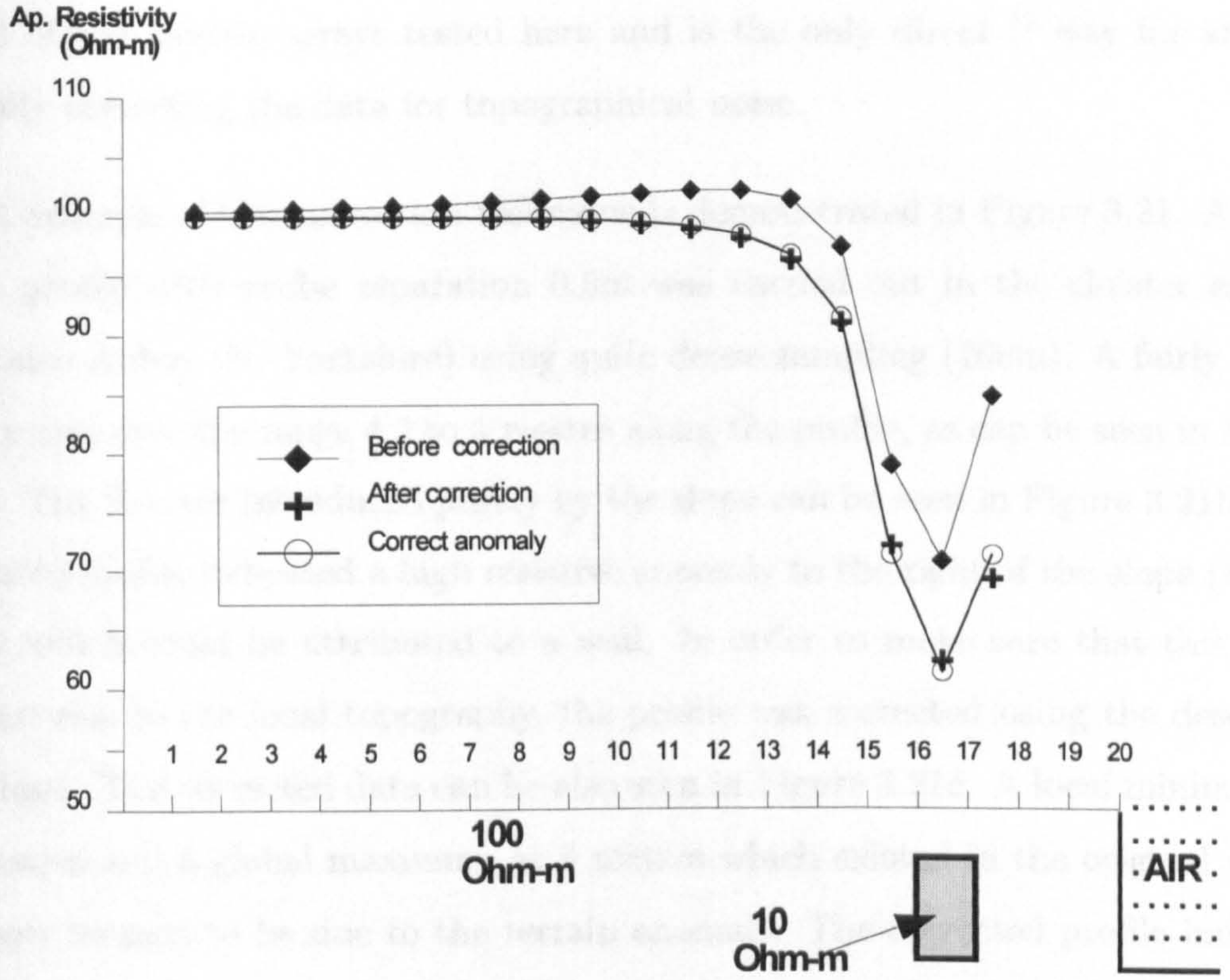


Figure 3.20: Correction of the topography effect for a pole-pole profile over a conductive prism (bottom).

- The corrected data set can now be interpreted as if there were no terrain variation.

An example of this procedure can be seen in Figure 3.20 for a case of a conductive prism under the of a vertical slope: The original pole-pole profile was corrected using the above procedure and the corrected profile is then compared with the anomaly that would have been obtained in the absence of the vertical discontinuity.

The corrected and "real" anomalies are in good agreement. Their misfit is explained if we consider the approximate nature of the correction scheme: the assumption being that the topography response may be superimposed to the response of the target. This is in contrast to the non-linear nature of the forward resistivity problem. However, despite this linearity assumption, the method produced satisfactory results

for all of the models/arrays tested here and is the only direct [12] way for at least partially correcting the data for topographical noise.

A real example of this correction technique is demonstrated in Figure 3.21. A twin-probe profile with probe separation 0.5m was carried out in the cloister area of Fountains Abbey (N. Yorkshire) using quite dense sampling (10cm). A fairly linear slope exists over the range 4.2 to 5 metres along the profile, as can be seen in Figure 3.21a. The % error introduced purely by the slope can be seen in Figure 3.21b. The measured profile indicated a high resistive anomaly to the right of the slope (Figure 3.21c) which could be attributed to a wall. In order to make sure that this effect was not due to the local topography, the profile was corrected using the described technique. The corrected data can be also seen in Figure 3.21c. A local minimum at 4.2 metres and a global maximum at 5 metres which existed in the original profile can now be seen to be due to the terrain anomaly. The corrected profile indicates the existence of a resistive feature which has a centre at approximately 5.4 metres.

3.5.4 Conclusions

Overall, the results confirm that topographical variations can have a significant impact on the field resistivity data values for all resistivity arrays and show that the significance of the distortions will vary according to the details of the topography and survey type. Topography effects are predictable and should be taken into account when designing surveys and when interpreting data.

Most importantly, it is shown that topographical variations can be treated in a flexible and accurate fashion within a realistic resistivity forward modelling process, pointing towards the viability of carrying out corrected resistivity data inversions/interpretations in the context of known topographical survey.

[12] Tong and Yang (1990) proposed an indirect technique which treats topographic corrections within the interpretation/inversion procedure.

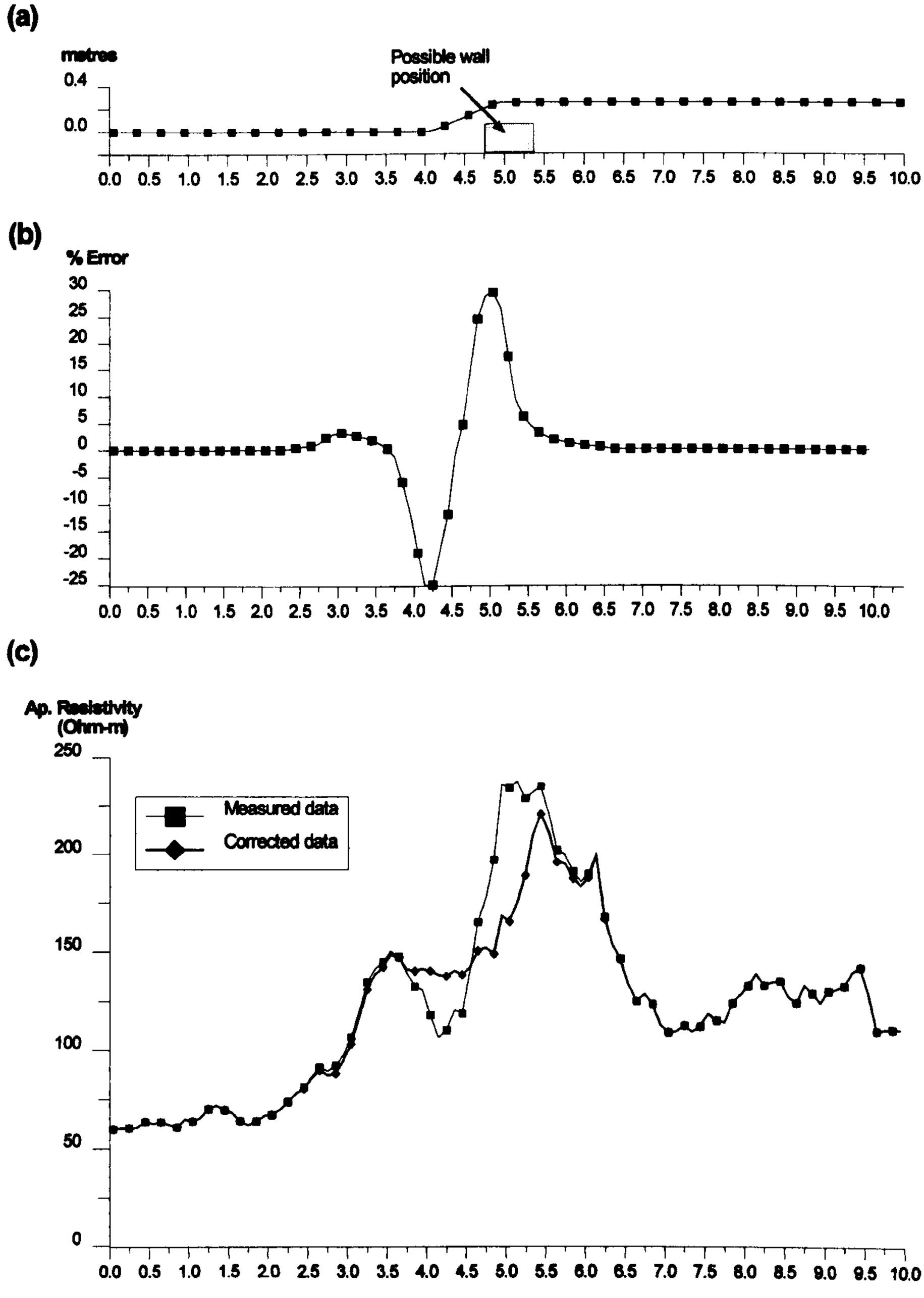


Figure 3.21: Correction of the topography effect for a twin-probe profile in the cloister of the Fountains Abbey (N. Yorkshire): a) the slope feature, b) % theoretical error introduced by the terrain c) the measured and corrected data sets.

The results of this study can be summarized as follows:

- The pole-pole array is more sensitive to the vertical slope topography for standard profiling applications while, conversely, the square-array is the least sensitive.

- For full 2-D data sets the dipole-dipole and pole-pole arrays produced anomalies of similar patterns. The Wenner array produced anomalies with opposite patterns. The pole-dipole array in general produced anomalies with asymmetric patterns.

- For 2-D surveys, as far as the amplitude is concerned, dipole-dipole and pole-dipole arrays produced the highest and lowest errors when compared with pole-pole and Wenner arrays. However, direct quantitative comparisons cannot be made since the sensitivity of each array is different (i.e. dipole-dipole may be giving the highest error, but it would also have given a higher and sharper anomaly under the presence of a target).

- All arrays produce significant, and potentially misleading, artificial errors within the data for slope angles larger than 10 degrees (if the extend of the slope is larger than the array spacing).

3.6 Chapter Overview

In this chapter the forward resistivity problem was addressed. The finite element method was chosen to simulate the earth resistivity responses due to its flexibility in modelling irregular boundaries. Further, modelling in 2.5 dimensions was found to be a reasonable compromise between accuracy and computational loads.

In the next part of this chapter the full theoretical development of a FEM 2.5-D earth resistivity modelling scheme was presented. The practical implementation of the scheme is discussed and its accuracy was tested.

Further, a strategy for modelling arrays parallel to the strike direction (i.e. square array) was proposed. The proposed technique modified an existing integration scheme and made it more flexible for modelling profiles of resistivity arrays which so far could not be modelled in 2.5 dimensions. The accuracy of the scheme was checked and results for various arrays were presented.

Finally, the FEM scheme was used to study the effects of terrain topography on commonly used resistivity arrays (dipole-dipole, Wenner, pole-pole, pole-dipole and Square). A mesh-generating algorithm was used to allow the realistic representation of the terrain topography. Different topographical terrain contexts, such as valleys, hills and steep slopes were tested. Topography effects are predictable and should be taken into account when designing surveys and when interpreting data.

The results confirm that topographical variations can have a significant impact on the field resistivity data values for all resistivity arrays. Further, it is shown that topographical variations can be treated in a flexible and accurate fashion within a realistic resistivity forward modelling process.

Chapter 4

The 2-D Inversion Procedure - Calculation of The Jacobian Matrix

In the first part of this chapter the general concept of resistivity inversion is presented. The theoretical and practical limitations of the inversion procedure are discussed and general approximate and accurate ways for treating the resistivity inverse problem are presented.

It will be shown that the Jacobian matrix is crucial in both approximate and accurate schemes. Therefore, before attempting to describe the inversion schemes analytically (this will be covered in the following two chapters), ways for calculating the Jacobian matrix are presented in detail. Particular emphasis is given to the application of those techniques within the finite element framework.

One of those techniques is used to calculate the sensitivity of commonly used arrays. It will be shown that the sensitivity matrix can be used to justify the response of the arrays and to design optimum resistivity surveys.

4.1 The Inversion Procedure

So far, we have been dealing extensively with the solution of the forward resistivity problem: given a known resistivity, calculate the observed data set. The inverse procedure (as the name suggests) is to find the subsurface distribution, given the observed data set.

Since we are dealing with discrete space, the subsurface is divided into a finite number of continuous blocks. (see Figure 4.1b) The resistivity within each block is considered to be constant. Note that each block is assumed to extend infinitely along the strike (y) direction: in other words, the resistivity distribution is treated as pseudo-two-dimensional (see Figure 4.1a). This is why the inversion is described as 2-D.

The overall resistivity distribution can be represented by a vector $\mathbf{x}$. Suppose that the vector $\mathbf{d}$ represents the measurements $\mathbf{d} = \{\mathbf{d_1}, \mathbf{d_2}, ...\mathbf{d_n}\}$ corresponding to that resistivity distribution. The forward modelling problem is to find a transformation $\mathbf{T}$ which connects the known resistivity vector $\mathbf{x}$ to the unknown measurement vector.

$$\mathbf{d} = \mathbf{T}(\mathbf{x}) \tag{4.1}$$

The transformation T is related to the solution of the Poisson's equation and here it was handled by using the finite element method. The inverse problem is to find an inverse transformation $\mathbf{T^{-1}}$ which connects the known vector of the measurements $\mathbf{d}$ with the unknown resistivity $\mathbf{x}$, that is

$$\mathbf{x} = \mathbf{T^{-1}}(\mathbf{d}) \tag{4.2}$$

Is inversion possible? One of the first questions to be answered is whether the inversion of resistivity data is possible: that is, whether the measured data set contains enough information to allow the resistivity to be determined uniquely, or (equivalently) to establish whether different resistivity distributions can always be distinguished by boundary measurements.

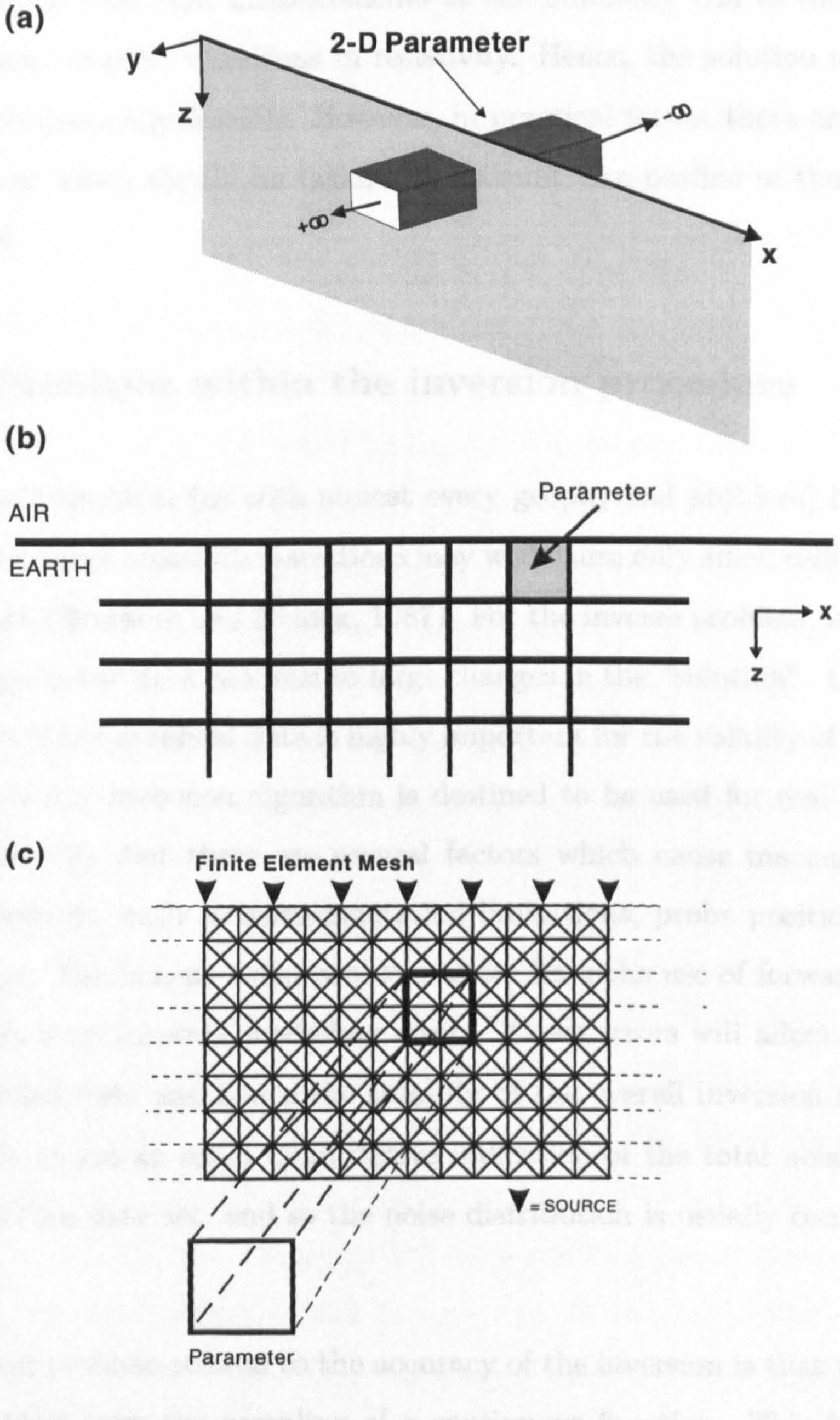


Figure 4.1: a) A 2-D parameter. b) The parametrization procedure. c) A parameter within the FEM scheme.

Recent studies [i.e. Kohn and Vogelious (1984)], involving rather complicated theoretical mathematical formulations (the presentation of which is beyond the scope of this work), proved that measurements at the boundary can in theory uniquely determine even smooth variations in resistivity. Hence, the solution of the inverse problem is theoretically feasible. However, in practical terms, there are still several considerations which should be taken into account. An outline of these considerations follows.

4.1.1 Problems within the inversion procedure

The resistivity problem (as with almost every geophysical problem) is ill-posed in the sense that large resistivity variations may well cause only small differences in the observed data (Breckom and Pidock, 1987). For the inverse problem, it follows that small changes in the data can lead to large changes in the "solution". Consequently, the accuracy of the observed data is highly important for the validity of the inversion results. Since any inversion algorithm is destined to be used for real data, we can recall (chapter 2) that there are several factors which cause inaccuracies within the measurements, such as instrumentation limitations, probe positioning, terrain anomalies etc. Further, an extra problem arises from the use of forward modelling, which suffers from inherent modelling errors. These errors will affect the accuracy of the modelled data and hence the stability of the overall inversion scheme. It is very difficult to get an exact quantitative indication of the total noise level which corresponds to a data set, and so the noise distribution is usually considered to be Gaussian.

An additional problem related to the accuracy of the inversion is that the measured data set derives from the sampling of a continuous function. This fact results in an unavoidable loss of information despite the fact that the sampling density can be chosen to be consistent with the size of the targets sought. Certainly, a high density measuring scheme has the potential to improve the quality of the obtained

information, and therefore the inversion results, but there are several restrictions imposed by practical difficulties. Automatic measuring systems are an obvious way of increasing the density of the measurements, but any inversion scheme would then have to deal with an increased amount of data computation- considering the current computer speed and memory, the extra computer power needed is not justified for most applications since the potential increase in the quality of the inversion results is not large enough compared to the cost (Smith and Vozoff, 1984).

The subsurface discretization poses another source of inaccuracy. In real terms the subsurface resistivity distribution is too complicated to be modelled in great detail by a set of discrete blocks. It could be argued that those blocks can be as small as desired but this is not the case: their number and size is limited by the type and amount of the observed data. Consequently the introduction of *a priori* parameter blocks generates artificial boundaries which are quite unlikely to coincide with the boundaries of the targets which we are trying to resolve. In this way the exact shape and properties of the target are impossible to reconstruct accurately. Of course, a high measurement density results in a correspondingly higher parameter density, and thus improved accuracy, but the limitations explained in the previous paragraph apply here as well.

The problems described so far are inherent to any inversion scheme. For the 2-D scheme an extra source of errors is the assumption that subsurface resistivity varies in only two dimensions. This is valid for a series of possible targets (i.e. layers, faults) and can be quite close to reality for several others (i.e. walls, ditches) and, in general, for any types of targets which have a lateral extent within the "range" of significant array sensitivity. But targets such as isolated voids can clearly be described accurately only by a 3-D resistivity variation[1]. An obvious solution to this problem is to use full 3-D schemes but the computational load of such a scheme

[1]Bear in mind that those modelling approximation errors are not due to the inversion scheme itself but are due to the 2.5-D forward modelling scheme which is used by the inversion in order to reproduce the original data. It will be shown in a later section that the inversion and forward modelling techniques are in most schemes strongly interrelated.

renders it impractical for routine data interpretation, at least for the time being.

The potential sources of errors in a resistivity inversion scheme can be summarized as follows:

- The ill-conditioned nature of the problem in conjunction with the limited accuracy of the observations.
- The fact that the measurements are just a discrete, usually sparse, sample of a continuous function.
- The subsurface resistivity is usually too complicated to be modelled accurately by a set of discreet blocks.
- The 2-D assumption is not valid for all possible targets.

Several methods have been suggested for handling some of these problems. They are based on using mechanisms for stabilizing the inversion procedure and on incorporating prior information about the area of interest and the quality of the measured data. Those techniques will be described.

Why invert? There are several factors that affect the performance of an inversion algorithm. Given all of these problems, questions about the practical usefulness of resistivity inversion may arise. The answer to these questions is that inversion should not be viewed as a panacea for solving the resistivity problem, but simply as a tool, which, when it is used properly, could be of great help for resistivity interpretation since (as it will be shown) direct interpretation of the resistivity data (e.g. via a pseudosection) is inadequate and possibly misleading. "Proper" use of the inversion algorithms means that the operator is aware of the limitations and the problems of the particular inversion scheme. Hence, an objective validation of the inversion results can be made.

The schemes for handling the 2-D resistivity inversion can be separated into two categories:

Approximate inversion schemes: including a wide range of algorithms which seek to simplify the inverse problem by assuming that it is linear.

Accurate inversion schemes: the inversion problem is handled as a non-linear procedure.

In the following chapters a detailed presentation of the algorithms used for 2-D reconstruction of the resistivity data is given. However, firstly it is necessary to present the general pattern of the suggested solutions (accurate and approximate) of the inversion problem.

4.2 General Solutions for the Resistivity Inverse Problem

Suppose that the resistivity distribution is represented by a vector $\mathbf{x}$ which has N parameters $\mathbf{x} = \{x_1, x_2 ... x_N\}$ and the respective M measured data by a vector $\mathbf{d}$. Note that the vector $\mathbf{d}$ could either represent apparent resistivity or potential difference measurements, since given apparent resistivity the potential difference can be easily calculated and vice-versa (recall equation 2.36). The set of algebraic equations that has to be solved is

$$\mathbf{f(x)} = \mathbf{d} \tag{4.3}$$

Due to the non-linearity of $\mathbf{f}$ the solution cannot be obtained by direct inversion.

Iterative non-linear schemes

The accurate inversion techniques are able to handle non-linearity by using an iterative optimization scheme. The basic pattern of such a scheme is to start with an initial resistivity model $\mathbf{x_o}$ and successively correct it in order for the modelled data $\mathbf{f(x_{current})}$ (which are the data that correspond to the current resistivity model) to be fitted to the measured data set $\mathbf{d}$. Since the partial derivatives of the modelled data with respect to the resistivity parameters can be obtained, (ways for calculat-

ing these derivatives are demonstrated in the next section) a Gauss method can be used. Expand $\mathbf{f(x)}$ with respect to a small resistivity variation dx in a Taylor series expansion

$$f(x_i + dx_i) = f(x_i) + \frac{\partial f(x_i)}{\partial x_i}\, dx_i \; + O(\; (dx_i)^2\;) \qquad i = 1, 2, ...N \qquad (4.4)$$

where $O((dx_i)^2)$ represents the higher order terms. Since dx is considered to be "small" the higher order terms can be neglected; hence equation 4.4 can be rewritten in a general form

$$\mathbf{f(x + dx)} \;=\; \mathbf{f(x)} \;+\; \mathbf{J dx} \qquad (4.5)$$

where **J** is the m x n matrix of the first derivatives (Jacobian matrix). The way that the Jacobian matrix can be derived will be shown in a following section.

Assume now that a criterion has been defined for the optimum solution of **dx** to be found from equation 4.5 given all other components. An iterative procedure can then be defined:

if $\mathbf{x^k}$ is the resistivity estimate at the k^{th} iteration

1 calculate the modelled data set $\mathbf{f(x^k)}$ using the forward model.

2 find the optimum $\mathbf{dx^k}$

3 set $\mathbf{x^{k+1} = x^k + dx^k}$

4 repeat until a criterion of fitness is satisfied.

Approximate schemes

The approximate schemes try to linearize the problem by assuming the following type of relationship

$$\mathbf{A\, x} \;\simeq\; \mathbf{d} \qquad (4.6)$$

Where **A** represents an approximate relationship between the parameters and the observations. The solution is then given by either directly inverting **A**, or, most

commonly, by approximating its inverse.

$$\mathbf{x} \simeq \mathbf{B}\ \mathbf{d} \tag{4.7}$$

where $\mathbf{B} \simeq \mathbf{A}^{-1}$. It has to be noted that iterative schemes can be also used.

4.2.1 Fitting criterion

We have explained that an inversion procedure is essentially an optimization scheme in which we are trying to fit the modelled data to the real data. In order to check the degree of fit the root mean square (RMS) error between the observed and calculated data is calculated. The RMS error is given by the following relationship (Jupp and Vozoff, 1975):

$$\text{RMS error} = \frac{1}{m}\sqrt{\sum_{i=1}^{m}\left(\frac{do_i - dc_i}{do_i}\right)^2} \tag{4.8}$$

Where do_i, dc_i is the i observed and calculated measurements respectively.

In an iterative scheme the RMS error can be also used as a stopping criterion: If the RMS error is not dropping consistently throughout each iteration, then no improvement in the fit is achieved and therefore the procedure can end. Similarly the procedure can end when the RMS error grows (divergence).

4.3 Calculation of the Jacobian Matrix

The inversion procedure involves the calculation of a resistivity distribution which will produce a modelled data set which is as close as possible to the observed one. The resistivity distribution in terms of the model space is expressed as a set of homogeneous blocks (parameters) which are allowed to vary their resistivity independently In most inversion schemes it is necessary to use a matrix, the Jacobian **J**, which associates variations in the property of those parameters with variations

in the observed data. Before we proceed to investigate the inversion schemes more closely, it is essential to demonstrate ways for calculating the Jacobian matrix.

If n is the number of elements in the parameter vector, **p**, of the problem and m is the number of elements in the observed data vector, **d**, the Jacobian is a m x n matrix and its i,j element J_{ij} is given by:

$$J_{ij} = \frac{\partial d_i}{\partial \sigma_j} \tag{4.9}$$

Where σ_j is the conductivity of the parameter p_j The Jacobian matrix (JM) is also known as the sensitivity matrix since it represents the sensitivity of the measurements to small changes of the parameter property [2]. If the observed data are apparent resistivities, equation 4.9 can be written as

$$J_{ij} = \frac{\partial \rho_{ai}}{\partial \rho_j} = -\frac{\partial \rho_{ai}}{\partial \sigma_j} = \frac{2\pi}{GI}\frac{\partial \Delta V_i}{\partial \sigma_j} \tag{4.10}$$

In equation 4.10 the geometrical factor (G) and the intensity of the inserted current (I) is known,thus the JM can be obtained by finding the variation of the potential drop versus the variation of the conductivity.

In general, the JM can be calculated using the following three techniques (McGillivray and Oldenburg, 1990): **a)** sensitivity technique, **b)** adjoint equation technique, **c)** perturbation technique.

The calculation of the JM is strongly related to the forward modelling technique used. Hence, for the purposes of this work, the ways for calculating the JM will be described analytically within the context of the finite element method (FEM). It has to be stressed that when the FEM forward modelling scheme is used each parameter usually consists of several sub-elements (see Figure 4.1c).

---

[2]This type of derivative is also referred as Fréchet derivative.

4.3.1 Sensitivity technique

The sensitivity technique (ST) is widely used for the calculation of the JM (Sasaki 1982; Smith and Vozoff, 1984). It has quite a simple mathematical formulation and can be easily applied within differential modelling schemes such as the FEM and FDM.

Recall the final system of equations derived from the assembly of the element equations, which has the general form (equation 3.38).

$$\mathbf{K} \cdot \mathbf{A} = \mathbf{F}$$

where: $\mathbf{K}$ is the transformed stiffness matrix, $\mathbf{A}$ is the transformed nodal potential vector and $\mathbf{F}$ is the loading terms vector.

Partial differentiation of the global FEM equation with respect to the conductivity σ_p of a parameter p yields:

$$\frac{\partial}{\partial \sigma_p} (\mathbf{K} \cdot \mathbf{A}) = \frac{\partial \mathbf{F}}{\partial \sigma_p} \tag{4.11}$$

and application of the chain rule (the load vector $\mathbf{F}$ is independent of the resistivity of the parameters) yields:

$$\mathbf{K}\frac{\partial \mathbf{A}}{\partial \sigma_p} = -\mathbf{A}\frac{\partial \mathbf{K}}{\partial \sigma_p} \tag{4.12}$$

The terms of equation 4.12 can be easily calculated. The transformed stiffness matrix $\mathbf{K}$ and the transformed nodal potential vector $\mathbf{A}$ are already known since they are obtained from the solution of the forward problem.

The term $\partial \mathbf{K}/\partial \sigma_p$ can be found by direct differentiation of the elementary stiffness terms $\mathrm{K}^{(e)}$. Suppose that p is the parameter that is differentiated with conductivity σ_p. Since the elementary stiffness term $K_{ij}^{(e)}$ is a function of conductivity (recall equation 3.20), differentiation with respect to σ_p yields

$$\frac{\partial K_{ij}^{(e)}}{\partial \sigma_p} = \begin{cases} K_{ij}^{(e)} \sigma_p & \text{if } e \in p_j \\ 0 & \text{otherwise} \end{cases} \tag{4.13}$$

The matrix $\partial \mathbf{K}/\partial \sigma_p$ is formulated by assembling the elementary differentiated stiffness terms, in a fashion similar to the formation of stiffness matrix for the FEM.

Since the final system of equations which has to be solved (equation 4.12) is entirely similar to that solved at the forward problem, Gaussian elimination can also be used. Note that the forward reduction of the stiffness terms $\mathbf{K}$ has already occurred in the solution of the forward problem, thus only back-substitution is needed. The solution gives the partial derivatives of the transformed potential with respect to conductivity changes of each particular parameter. After solving the system of equation 4.12 for every wavenumber, the partial derivatives of the total potential are recovered by applying the inverse cosine Fourier transformation.

It should be noted that the pattern of the derivatives of the transformed potential is somewhat irregular: the derivative of the transformed potential for the first wave-number can sometimes be slightly smaller than the derivative of the transformed potential for the second wavenumber[3]. Hence, integration schemes such as subsectional exponential fitting will not suffice. Instead, for the area between (0,1], Gaussian quadrature was applied (Marron, 1982) using three critical wavenumber values. The quadrature weighting factors were decided using standard quadrature tables (Abramovitz and Stegun, 1972). Exponential fitting was then performed for the remaining wave-numbers.

As long as the derivatives of the nodal potential are known, then the point to point variation $\partial \Delta V/\partial \sigma$ can be easily calculated. Finally, the entries of the Jacobian matrix can be calculated using equation 4.10.

4.3.2 The adjoint equation technique

The adjoint equation technique (AET) has been applied both in 2.5-D and 3-D resistivity problems (McGillivray and Oldenburg 1990; Park and Van, 1991) but the

[3]This could be due to round-off errors

formulation was restricted to the finite difference method. Similar formulations for calculating the JM have been derived within the network analysis context (Trip et al. 1994; Yorkey 1987a). Since no application of the method to the 2.5-D FEM modelling scheme has been reported, in this section the equations for applying the AET into the FEM are derived.

The AET is based on the use of the properties of the adjoint operator and the adjoint Green's function (Lanczos, 1960). After laborious mathematical formulations (the presentation of which is beyond the scope of this work) it can be shown that for the resistivity case the sensitivity is a function of the differentiated original equation and the adjoint Green's function. Since the solution of the resistivity problem is self-adjoint, the same forward solver can be used to give the solution of both the original and the adjoint functions. McGillivray and Oldenburg (1990) derived the general form of the sensitivity of the corresponding transformed potential for the 2.5-D resistivity case:

$$\frac{\partial V(x_o,k,z_o,x_s,z_s)}{\partial \sigma_j} = \int_D \left[k^2\psi_j V(x,k,z,x_s,z_s) - \psi_j \frac{\partial^2 V(x,k,z,x_s,z_s)}{\partial^2 x} - \psi_j \frac{\partial^2 V(x,k,z,x_s,z_s)}{\partial^2 z}\right] \cdot V(x,k,z,x_o,z_o)\, dxdz \tag{4.14}$$

where:

- $\partial V(x_o,k,z_o,x_s,z_s)/\partial\sigma_j$ is the partial derivative of the transformed potential measurement at (x_o,k,z_o) when the source is at (x_s,z_s) in respect of parameter j.
- $V(x,k,z,x_s,z_s)$, $V(x,k,z,x_o,z_o)$ are the transformed potential at (x,k,z) when the source is at (x_s,z_s) and (x_o,z_o) respectively.
- ψ_j is 1 if the (x,z) coordinates are within the limits of parameter region j, and otherwise is 0.

Using a simplified notation [V(x_o,k,z_o,x_s,z_s)=V_o, V(x,k,z,x_s,z_s)=V, V(x,k,z,x_o,z_o)=V'] equation 4.14 can be rewritten as

$$\frac{\partial V_o}{\partial \sigma_j} = \int_D \left[k^2 \psi_j \; VV' - \psi_j \frac{\partial^2 V}{\partial^2 x} V' - \psi_j \frac{\partial^2 V}{\partial^2 z} V' \right] \; dxdz \tag{4.15}$$

By using the chain rule of differentiation equation 4.15 becomes:

$$\begin{aligned} \frac{\partial V_o}{\partial \sigma_j} = \int_D & \left[k^2 \psi_j \; VV' + \psi_j \frac{\partial V}{\partial x} \frac{\partial V'}{\partial x} + \psi_j \frac{\partial V}{\partial z} \frac{\partial V'}{\partial z} \right] \; dxdz \\ -\psi_j & \int_D \left[\frac{\partial}{\partial x} \left(\frac{\partial V}{\partial x} V' \right) + \frac{\partial}{\partial z} \left(\frac{\partial V}{\partial z} V' \right) \right] \; dxdz \end{aligned} \tag{4.16}$$

The last integral of the right hand side of equation 4.16 is a perfect differential and by using Green's theorem can be expressed as a line integral over the boundary B of the domain:

$$\int_D \left[\frac{\partial}{\partial x} \left(\frac{\partial V}{\partial x} V' \right) + \frac{\partial}{\partial z} \left(\frac{\partial V}{\partial z} V' \right) \right] dxdz = \oint_B \left[\frac{\partial \tilde{V}}{\partial x} V' n_x \; + \; \sigma \frac{\partial \tilde{V}}{\partial z} \; V' \; n_z \right] \; ds \tag{4.17}$$

The n_x, n_z symbols are the x,z components of the outward unit n normal to the domain's boundary. If the boundary conditions are considered, the line integral of equation 4.17 is zero: the potential (V') at the side and bottom boundaries is zero as a result of the Dirichlet BC ($V_{D\infty} = \; 0$), and at the air/earth interface, the normal component of the potential derivative is zero. Therefore, the equation that gives the sensitivity of the transformed potential finally becomes:

$$\frac{\partial V_o}{\partial \sigma_j} = \int_D \left[k^2 \psi_j \; VV' + \psi_j \frac{\partial V}{\partial x} \frac{\partial V'}{\partial x} + \psi_j \frac{\partial V}{\partial z} \frac{\partial V'}{\partial z} \right] \; dxdz \tag{4.18}$$

By using equation 4.18 the problem can now be formulated within the FEM context. Suppose that the sensitivity of the transformed potential (at a node n due to a source at node m) with respect to a perturbation of the conductivity of element e is to be found (see Figure 4.2). Equation 4.18 becomes

$$\frac{\partial V_{nm}}{\partial \sigma^{(e)}} = \int_{\Delta^{(e)}} \left[k^2 V_n^{(e)} V_m^{(e)} + \frac{\partial V_n^{(e)}}{\partial x} \frac{\partial V_m^{(e)}}{\partial x} + \frac{\partial V_n^{(e)}}{\partial z} \frac{\partial V_m^{(e)}}{\partial z} \right] \; dxdz \tag{4.19}$$

where V_{nm} is the potential at n due to a source at m and $V_m^{(e)}$, $V_n^{(e)}$ is the potential at element e due to sources at m, n respectively.

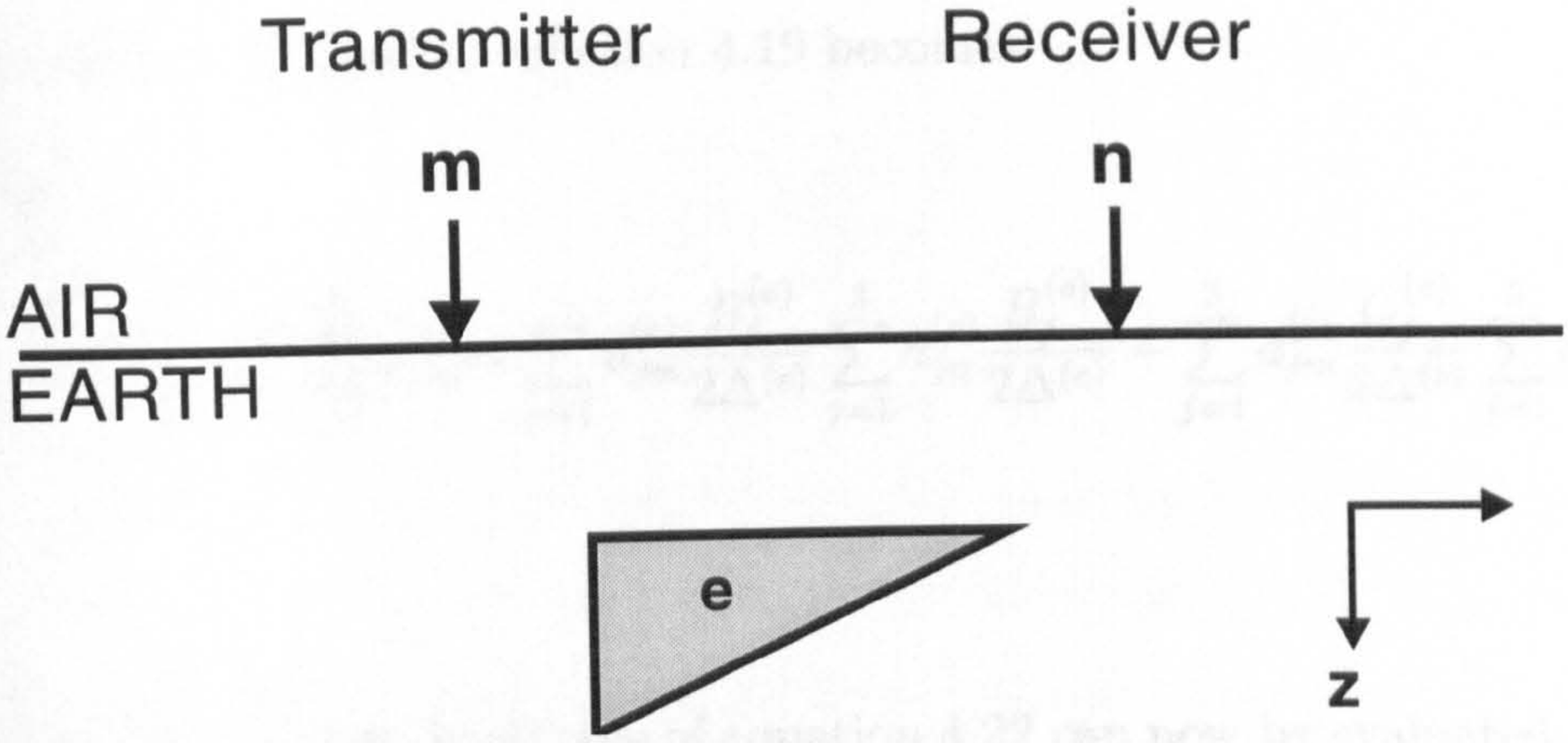


Figure 4.2: The receiver, transmitter, parameter configuration (see text).

The transformed potential derivatives in the right hand side of equation 4.19 can be calculated by the use of equation 3.26 which gives an element's transformed potential with respect to the potential at its nodes:

$$\tilde{V}^{(e)} = \sum_{j=1}^{3} a_j^{(e)} \phi_j^{(e)}$$

where

$$\phi_j^{(e)} = \frac{A_j^{(e)} + B_j^{(e)} x + C_j^{(e)} z}{2\Delta^{(e)}} \qquad j = 1, 2, 3$$

the terms $A_j^{(e)}, B_j^{(e)}, C_j^{(e)}$ depend on the coordinates of the node j, and $\Delta^{(e)}$ is the area of the element (see equation 3.28).

Therefore the derivatives of the right hand side of equation 4.19 will be

$$\frac{\partial \tilde{V}_m^{(e)}}{\partial x} = \sum_{j=1}^{3} a_{jm}^{(e)} \frac{\partial \phi_j^{(e)}}{\partial x} = \sum_{j=1}^{3} a_{jm}^{(e)} \frac{B_j^{(e)}}{2\Delta^{(e)}}$$

$$\frac{\partial \tilde{V}_m^{(e)}}{\partial z} = \sum_{j=1}^{3} a_{jm}^{(e)} \frac{\partial \phi_j^{(e)}}{\partial z} = \sum_{j=1}^{3} a_{jm}^{(e)} \frac{C_j^{(e)}}{2\Delta^{(e)}} \tag{4.20}$$

Similarly

$$\frac{\partial \tilde{V}_n^{(e)}}{\partial x} = \sum_{j=1}^{3} a_{jn}^{(e)} \frac{B_j^{(e)}}{2\Delta^{(e)}} \qquad \frac{\partial \tilde{V}_n^{(e)}}{\partial z} = \sum_{j=1}^{3} a_{jn}^{(e)} \frac{C_j^{(e)}}{2\Delta^{(e)}} \tag{4.21}$$

Due to equations 4.20,4.21, equation 4.19 becomes

$$\frac{\partial V_{nm}}{\partial \sigma^{(e)}} = \int_{\Delta^{(e)}} \left[k^2 \sum_{j=1}^{3} a_{jn}^{(e)} \sum_{j=1}^{3} a_{jm}^{(e)} + \sum_{j=1}^{3} a_{jm}^{(e)} \frac{B_j^{(e)}}{2\Delta^{(e)}} \sum_{j=1}^{3} a_{jn}^{(e)} \frac{B_j^{(e)}}{2\Delta^{(e)}} + \sum_{j=1}^{3} a_{jm}^{(e)} \frac{C_j^{(e)}}{2\Delta^{(e)}} \sum_{j=1}^{3} a_{jn}^{(e)} \frac{C_j^{(e)}}{2\Delta^{(e)}} \right] dxdz \tag{4.22}$$

The integral of the right hand side of equation 4.22 can now be evaluated. The first term can be evaluated using the triangle integration formula given in equation 3.32. The integral of the remaining two terms can be calculated directly. Equation 4.22 gets the following form

$$\begin{aligned} \frac{\partial V_{nm}}{\partial \sigma^{(e)}} &= K^2 \Delta^{(e)} \frac{\sum_{j=1}^{3} a_{jm}^{(e)} \sum_{j=1}^{3} a_{jn}^{(e)} + a_{1m}^{(e)} a_{1n}^{(e)} + a_{2m}^{(e)} a_{2n}^{(e)} + a_{3m}^{(e)} a_{3n}^{(e)}}{12} \\ &+ \frac{\sum_{j=1}^{3} a_{jm}^{(e)} B_j^{(e)} \sum_{j=1}^{3} a_{jn}^{(e)} B_j^{(e)} + \sum_{j=1}^{3} a_{jm}^{(e)} C_j^{(e)} \sum_{j=1}^{3} a_{jn}^{(e)} C_j^{(e)}}{4\Delta^{(e)}} \end{aligned} \tag{4.23}$$

Using equation 4.23 the sensitivity of any potential measurement due to perturbations of any element's conductivity can be calculated. It can be seen that the sensitivity is actually a function of the nodal coordinates and the nodal transformed potential of the element due to sources both at the transmitter and the receiver locations. The nodal potential due to sources at the transmitter is calculated anyway to get the forward response, thus only the potential due to sources at the transmitter needs to be evaluated. This may seem to suggest that the procedure is quite time consuming, but in a combined sounding profiling measuring scheme most of the electrodes which provide a potential measurement for one probe arrangement will be used to apply current in another. Therefore, only few extra sources positions need to be evaluated in order for all sensitivity combinations to be obtained (e.g. for the dipole-dipole scheme just two extra source positions have to be calculated).

As long as the derivatives of the transformed potential in respect to changes into the element's conductivity are known, the derivative with respect to a particular

parameter can be readily found:

$$\frac{\partial V}{\partial \sigma_p} = \sum_{i=1}^{q} \frac{\partial V}{\partial \sigma_i} \tag{4.24}$$

where i=1,2, ..q are the elements that form the parameter region (see Figure 4.1c).

Subsequently, the partial derivatives of the total potential are recovered by applying the inverse Fourier transformation in a fashion described into the previous section. Finally, the Jacobian matrix entries can be calculated using equation 4.10.

4.3.3 The perturbation technique

The perturbation technique (PT) gives an approximation to the Jacobian matrix in a finite difference sense. This method has been widely used in medical imaging (Kim and Woo., 1987; Yorkey, 1986) but it cannot strictly be classified as a method for obtaining the SM and hence the resultant matrix is called the Perturbation matrix (PM). The PM can give a useful indication of how every parameter affects the measurements.

The PM can be obtained in the following way: an initial resistivity distribution $\mathbf{p}$ is assumed and the measurement $\rho_i(\mathbf{p})$ which corresponds to this distribution is calculated. Subsequently, the resistivity of one parameter, say j, is perturbed by Δp_j and the new measurement $\rho_i(\mathbf{p} + \Delta p_j)$ is obtained. The ij element of the perturbation matrix $\mathbf{P}$ is then given by

$$P_{ij} = \frac{\rho_i(\mathbf{p} + \Delta p_j) - \rho_i(\mathbf{p})}{\Delta p_j} \approx J_{ij} \tag{4.25}$$

The approximate nature of the PM is due to the arbitrary selection of the perturbation quantity Δp_j (Broyden, 1970): If Δp_j is "small" it could cause excessive rounding error (substraction of two very similar numbers) and if it is "large" it violates the assumptions of the first derivative. Further, the PT is a particular time-consuming procedure since the calculation of the entire PM needs as many forward calculations as the existing parameters.

On the other hand the PT is attractive for its computing simplicity: the forward solver needs only minor modification to accommodate the technique.

4.3.4 Results

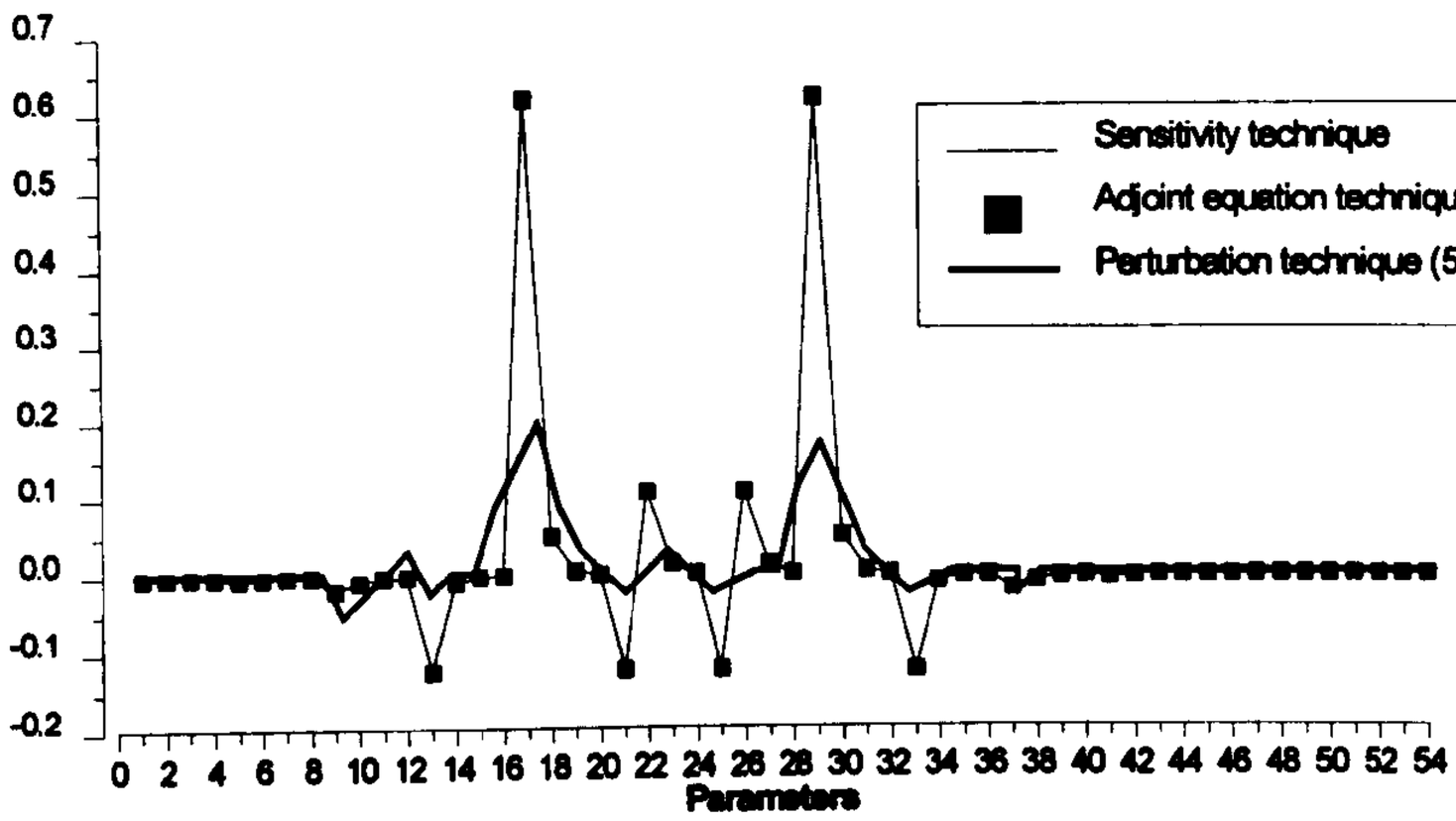
All three techniques have been incorporated within the FEM algorithm and tested. As far as the accuracy is concerned the sensitivity and the adjoint equation techniques produced identical results (Figure 4.3a). The results of the PT were not very close to these of the other two methods (Figure 4.3a).

In order to check the overall accuracy of the calculations, the computed results of the adjoint equation technique were compared to analytical results given by Sasaki (1982) for a dipole-dipole sounding over a three layered earth [see Figure 4.3b (top)]. Each layer is considered to be a parameter and graphs of the analytical and computed sensitivity entries for the first and second layers are depicted in Figure 4.3b. It can be seen that the analytical and computed responses are in good agreement.

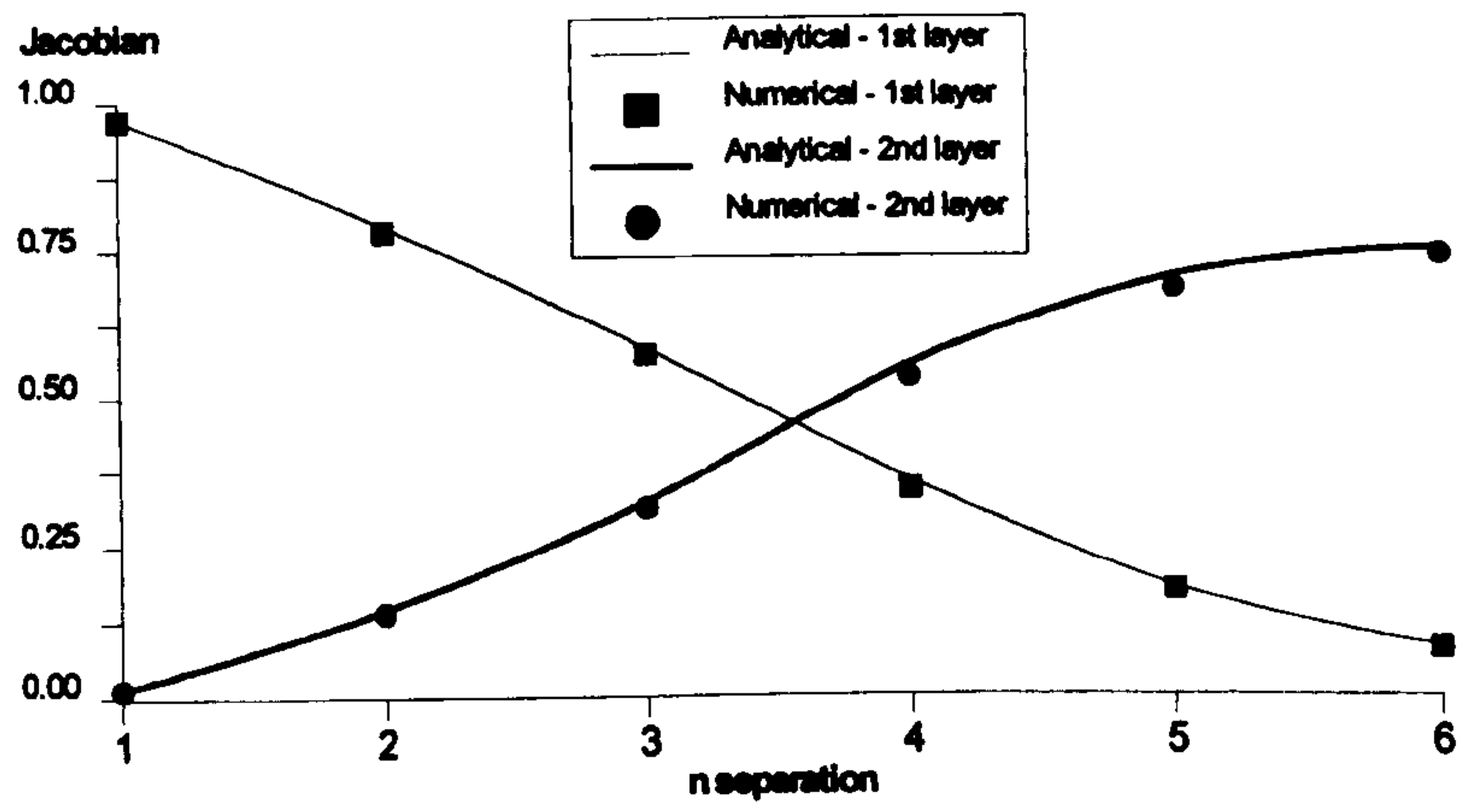
The computational time each method takes is quite important since the computation of the JM is quite a time-consuming procedure. The adjoint equation technique involves less floating point operations (Yorkey 1986) and thus was faster than the sensitivity technique. The perturbation technique proved to be the most time-consuming.

In conclusion, in this work the AET was used as the standard method for calculating the Jacobian matrix since it proved to be both accurate and economical in computational time.

(a)



(b)



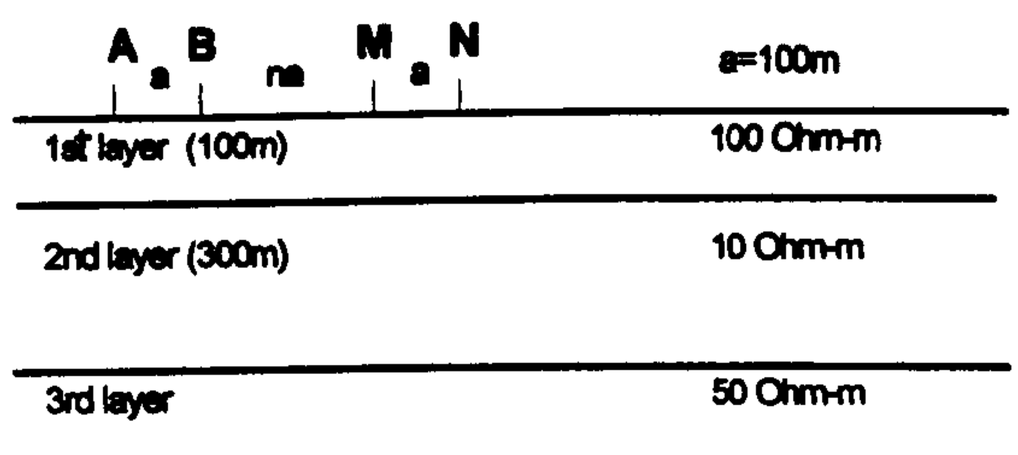


Figure 4.3: a) the JM calculated using different techniques, b) comparison between analytical and numerical (AET) results for a 3-layer model (bottom).

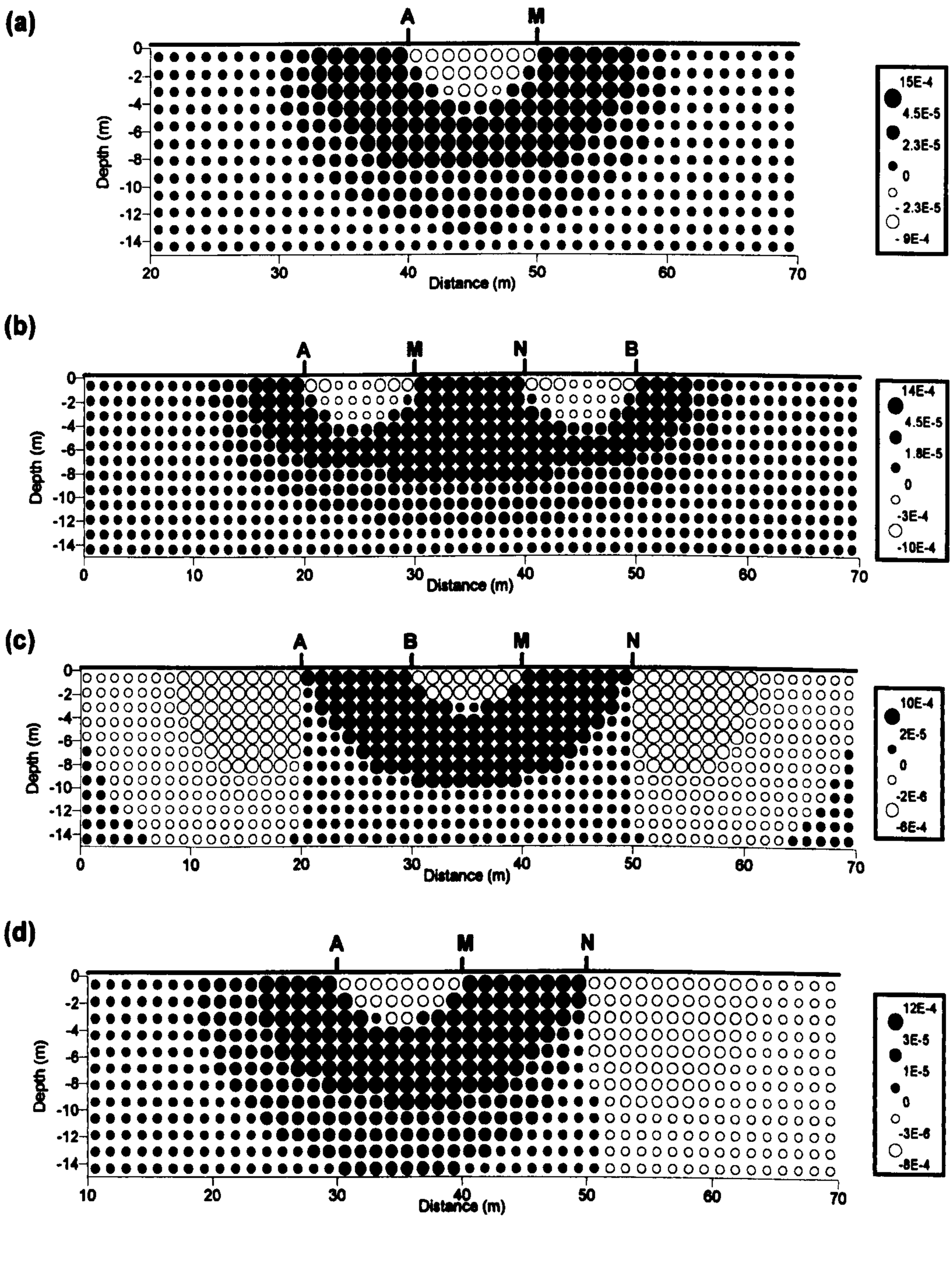


Figure 4.4: The variation of the JM for common probe arrangements (homogeneous ground): a) pole-pole, b) Wenner, c) dipole-dipole, d) pole-dipole.

4.4 The Variation of the JM for Common Probe Arrangements

The AET was used to calculate the sensitivity matrix responses for common resistivity arrays. A homogeneous ground divided into a set of square parameters was considered. In Figure 4.4 the sensitivity matrix entries for pole-pole, Wenner, dipole-dipole, and pole-dipole arrays are shown. The exact electrode location for each configuration is also depicted.

From the results it is clear that there are positive and negative JM entries. Positive JM values indicate that positive or negative perturbations of the conductivity will cause respectively positive or negative changes into the measurements. Conversely, negative JM values indicate that positive or negative perturbations of the conductivity will cause respectively negative or positive changes into the measurements.

The most important sensitivity to consider is that of the pole-pole array, since the sensitivities of the other arrays are a superposition of elementary pole-pole sensitivities. The pole-pole array has a high positive sensitivity area around the two measuring probes and a negative sensitivity area just below the region of the measuring probes. This negative area extends to a depth of approximately 0.3a (where a is the inter-probe spacing).

The sensitivity for the Wenner array is symmetrical. It appears as high positive values at the area between the potential probes and outside the current probes. High negative values appear at the regions between the potential and current probes and extend to a depth of approximately 0.3a.

The dipole-dipole array has high negative sensitivity areas between the two dipoles as well as left and right of them. High positive sensitivity regions occur below the high negative sensitivity region but the highest values appear at the regions below the current and potential dipoles. Finally the pole-dipole, produces an asymmetric

sensitivity distribution with high negative sensitivities just below the centre of the array and to the right of the potential dipole. The positive sensitivity regions are located in the regions just below the potential dipole and below the central negative sensitivity region, as well as to the right of the current probe.

No direct quantitative conclusions about the field measurements can be drawn from the Jacobian matrix since a measurement is the integrated effect of the individual parameters, but it can be said that for all arrays the positive entries are predominant and appear to have a larger amplitude. This explains the fact (which is the basic assumption when interpreting resistivity data) that, broadly speaking, a typical array response to a resistive (or conductive) body produces a predominantly positive (or negative) anomaly. However the detailed variations of the Jacobian matrix for the various arrays makes it clear that this is not always the case. In fact, the Jacobian matrix can give a useful qualitative insight into the response of resistivity arrays to certain features (i.e. the asymmetric pattern of pole-dipole anomalies is due to the asymmetric nature of the sensitivities of that array) and explain complicated and potentially misleading anomaly patterns which are contradictory to the common assumptions of resistivity surveying. In the following example illustrations of such cases are presented.

The responses of the pole-pole, Wenner and dipole-dipole arrays over a thin resistive body buried at a depth of 0.25a (where a is the inter-probe spacing) are presented in Figure 4.5. The configuration and initial positioning for each array can be seen at the bottom of the x-axis of each graph. The results can be explained easily by referring to the sensitivity distribution of each array (see Figure 4.4). The pole-pole array (Figure 4.5b) produces a negative anomaly when it is over the body due to the negative sensitivity occurring in its upper central region. Similarly, the Wenner array (Figure 4.5c) produces two negative anomalies on either side of the body which are due to the negative sensitivity of the areas between the current and potential probes. Finally, the two peak anomaly produced by the dipole-dipole array (Figure 4.5d) is due to the high sensitivity of the array below the potential and current

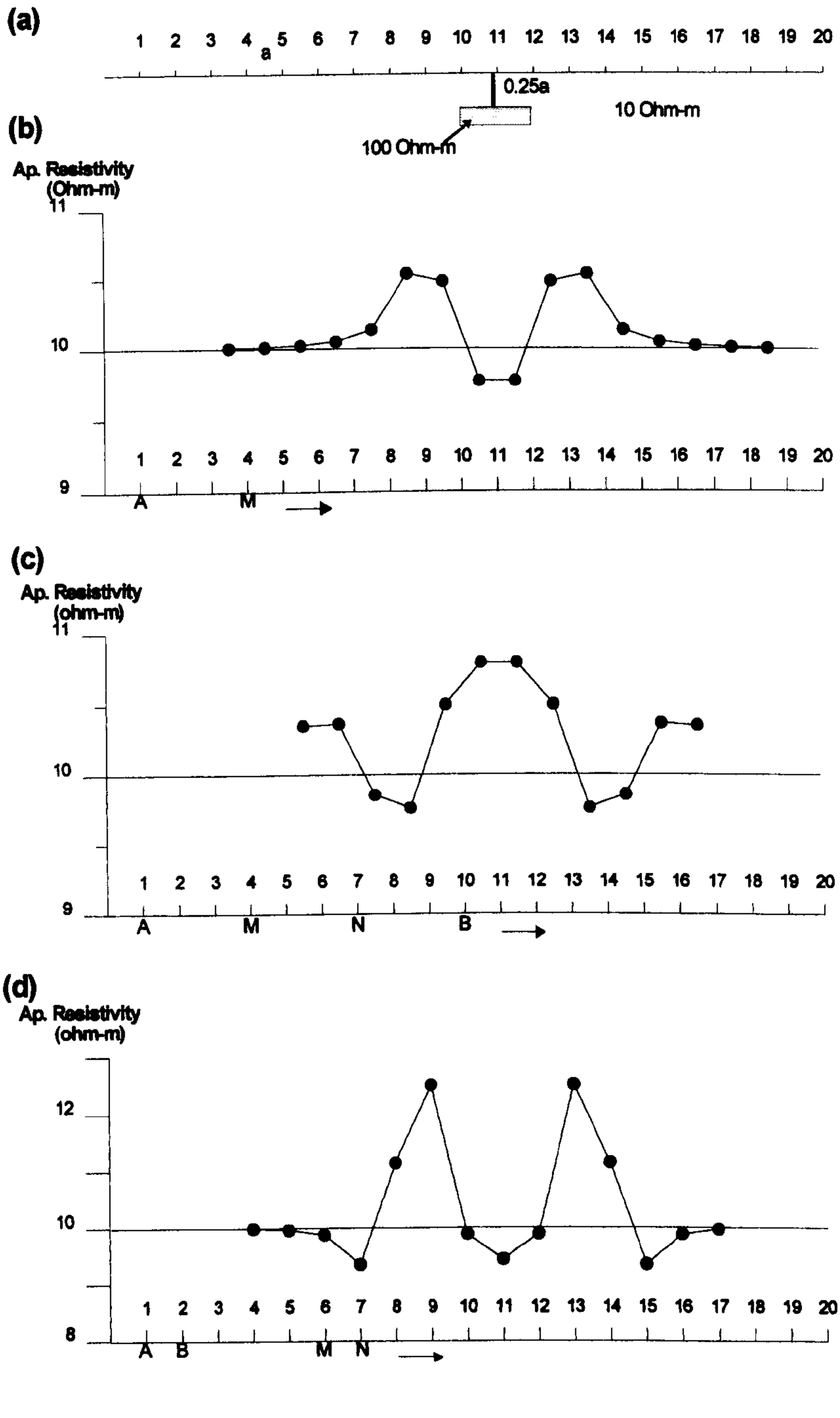


Figure 4.5: The response of various arrays to a shallow resistive prism: a) the model, b) pole-pole, c) Wenner, d) dipole-dipole.

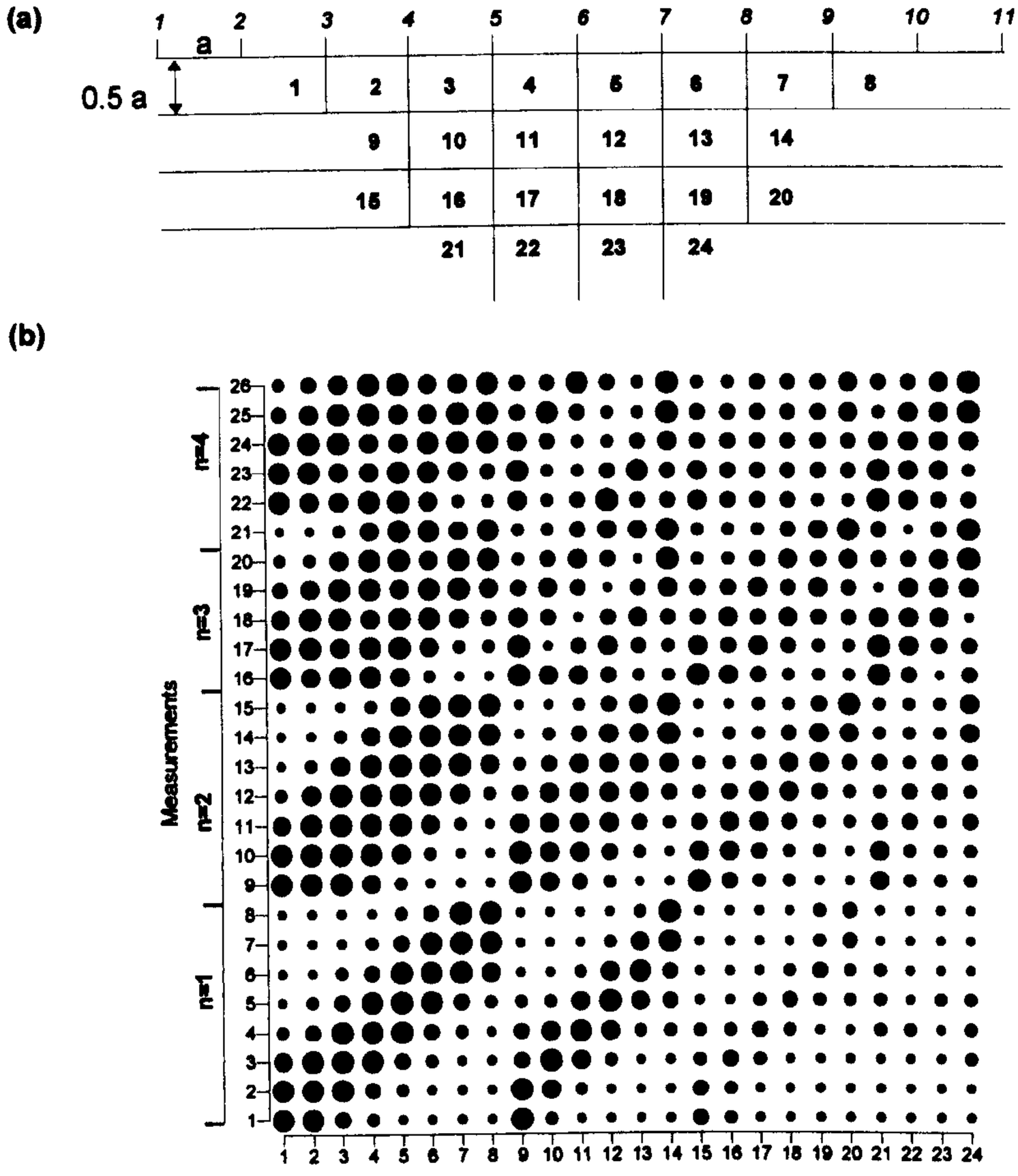


Figure 4.6: a) The parametrized space, b) the entries of the Jacobian matrix.

dipoles.

4.4.1 Sensitivity analysis

Another observation is that the amplitudes of the JM values vary abruptly between parameter regions: there are parameters which exert a large influence on the measurement and others that have only a small impact (almost zero). The latter are considered to be "unimportant" since they cannot convey any significant information about the subsurface to that particular measurement, and therefore they cannot be well-resolved by this measurement - even a very large change in the resistivity of that parameter will have a negligible effect on the measurement.

Based on this observation, an overall evaluation of a data set with respect to the parameters of the subsurface can be made. For a particular set of data and parameters the Jacobian matrix is calculated. If there is a column (say j) of the Jacobian matrix which has considerably small sensitivity values then the particular data set cannot resolve the j parameter.

This procedure can be used to design an optimum resistivity survey for a particular site. In practical terms, when surveying, the basic characteristics of the possible targets (approximate depth/size) as well as the area of interest are often known *a priori*, so that the subsurface can be parametrized accordingly. Subsequently, by a trial and error Jacobian testing the measured data set can be designed so than there are no "unimportant" parameters. On the other hand, if the measured data have already been obtained, the technique can be used to achieve an optimum parametrization of the area of interest.

To give an example, in Figure 4.6a the parametrization corresponding to a dipole-dipole data set with n=4 can be seen. The resultant Jacobian matrix was obtained and the "post-map" of the absolute value of its elements can be seen in Figure 4.6b. It can be seen that there are no unimportant parameters for this example.

In order to check the conditioning of the resultant Jacobian matrix its eigenvalues were obtained using the Singular Value Decomposition (SVD) technique (Press et al, 1987) [4]. The eigenvalues of the matrix can be shown in Figure 4.7. The condition number of the matrix (the ratio between the largest and smallest eigenvalues) was found to be 1000, and indication that the matrix is ill conditioned. The ill-conditioning of the Jacobian matrix is unavoidable for resistivity inversion. In a following chapter it will be explained how ill-conditioning is related to the physics of the problem as well as to the accuracy of the observations.

---

[4]Using the SVD technique (Lanczos, 1960) a matrix **A** can be decomposed into three matrices: **A=U W V**, where **U** and **V** are orthogonal matrices and **W** is a diagonal matrix which contains the eigenvalues of **A**. Further applications of the technique are presented in a following chapter.

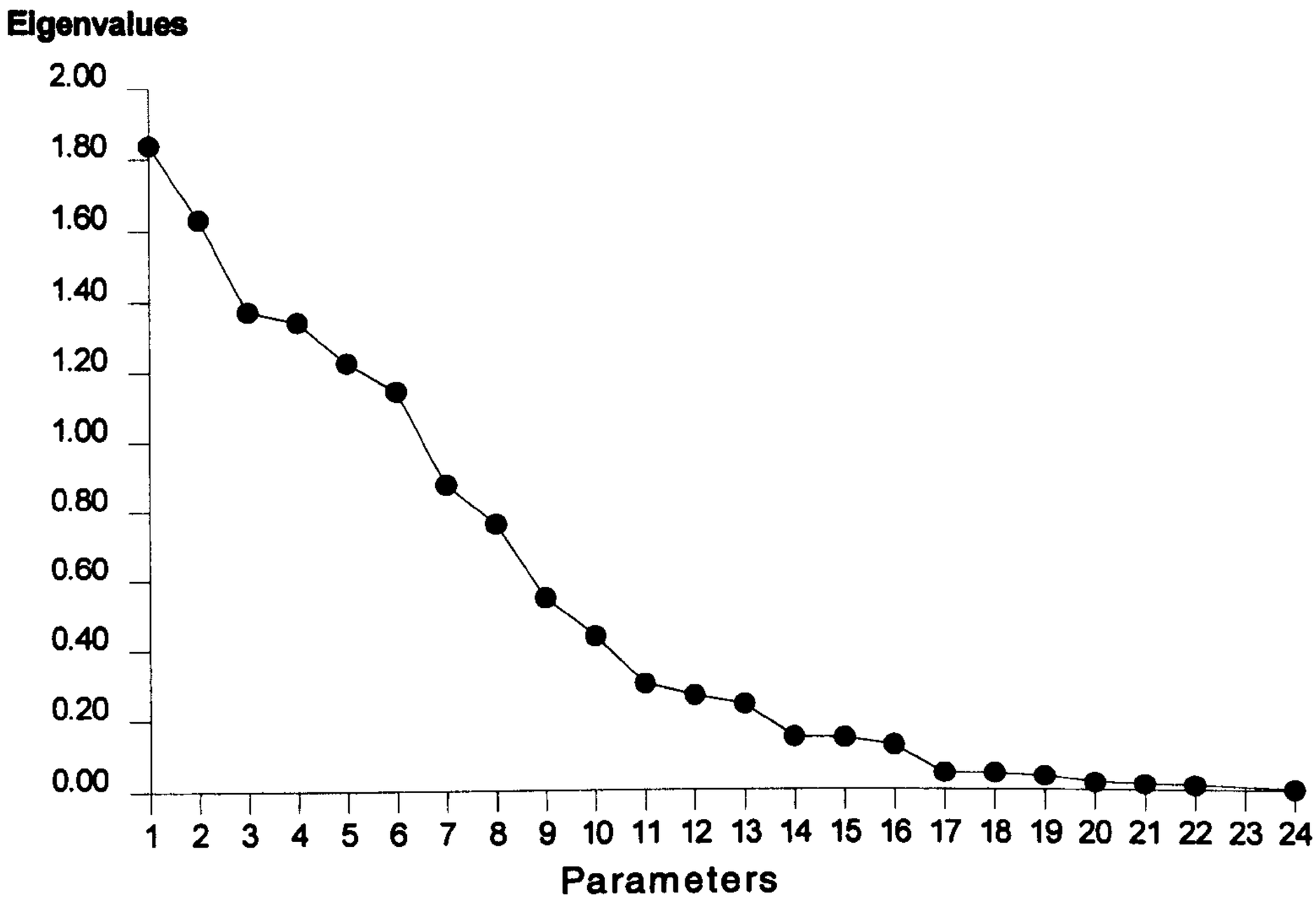


Figure 4.7: The eigenvalues of the Jacobian matrix.

4.5 Chapter Overview

In this chapter the general principles of the resistivity inversion problem have been described. There are several factors that can render 2-D inversion inaccurate: the ill-conditioned nature of the problem, data inaccuracies, the discrete nature of the measurements and parameters, and the 2-D assumption for the variations in the resistivity distribution.

Despite these problems the inversion procedure is absolutely worthwhile, since it is the only way to get additional information from the data. The known resistivity inversion schemes can be separated into two categories: accurate inversion schemes which treat the problem as fully non-linear, and approximate inversion schemes, which seek to solve the inverse problem by assuming that it is linear. The general concepts of those categories were explained. Furthermore, it was shown that an important factor within the inversion procedure is the Jacobian matrix.

In the second part of this chapter a detailed description of three techniques for cal-

culating the Jacobian matrix within the FEM context was given. The equations for applying these techniques were developed explicitly. The adjoint equation technique was found to be accurate and quite economical in computing time.

The adjoint equation technique was then used to calculate the sensitivity response of common resistivity arrays. The results indicate that the Jacobian matrix can be used to get a qualitative analysis of the response of the arrays over known targets. It was shown that unusual apparent resistivity responses can be explained adequately. Finally, the use of the Jacobian matrix as a means for designing optimum surveys or achieving optimum parametrization was demonstrated.

Chapter 5

2-D Reconstruction of Earth Resistivity Data using Approximate Inversion Schemes

In this chapter commonly used approximate inversion algorithms such as the pseudo-section, the Zhody-Barker method and the back-projection technique will be reviewed. Relevant medical imaging algorithms will be presented as well. The advantages and disadvantages of the techniques will be discussed and several examples with synthetic data will be presented. An explanation of why these approximate techniques are effective is given by means of the Jacobian matrix.

In the second part of this chapter a generalized (iterative) back-projection algorithm is proposed. The need for such an algorithm derives from an analysis of the limitations of the existing approximate algorithms. The algorithm can include all of the presented back-projection algorithms, as well as techniques such as the pseudo-section and Barker's method. Further, it can reconstruct data from any array as well as from unconventional probe arrangements. The merits and limitations of the algorithm are discussed and examples of its performance with synthetic and real data are given.

5.1 A Review of Approximate Reconstruction Algorithms

The "accurate" inversion schemes can give good quality results, but these algorithms are quite costly in terms of computer time and memory. An alternative strategy towards resistivity inversion is to use approximate schemes, which, despite their intrinsic theoretical weaknesses, can produce reasonably valid sectional images of the subsurface resistivity pattern in a fairly short time. The need for such fast algorithms is increased by the development of automatic measuring systems which can result in a huge increase in the number of recorded measurements.

Approximate algorithms handle the inversion problem as a linear one and they serve as a tool for preliminary interpretation of the experimental data sets. For most approximate resistivity reconstruction techniques it is only physical intuition rather than a solid mathematical background which is used to justify why they do work. Furthermore, it is quite possible to produce results which can lead to erroneous interpretations, and hence a thorough study of their properties and limitations is needed.

In this section several approximate algorithms from the geophysical field will be reviewed, but first a brief description of related medical algorithms is presented. This is necessary because most of the recent geophysical back-projection techniques are based on medical back-projection algorithms.

Tested models

Most of the techniques presented in this section were tested using a series of synthetic data sets. Four of the models which were used to produce the synthetic data are depicted in Figure 5.1. We will be referring to those models as Model 1 (one resistive prism), Model 2 (vertical discontinuity), Model 3 (two resistive prisms), Model 4 (conductive layers and a resistive prism).

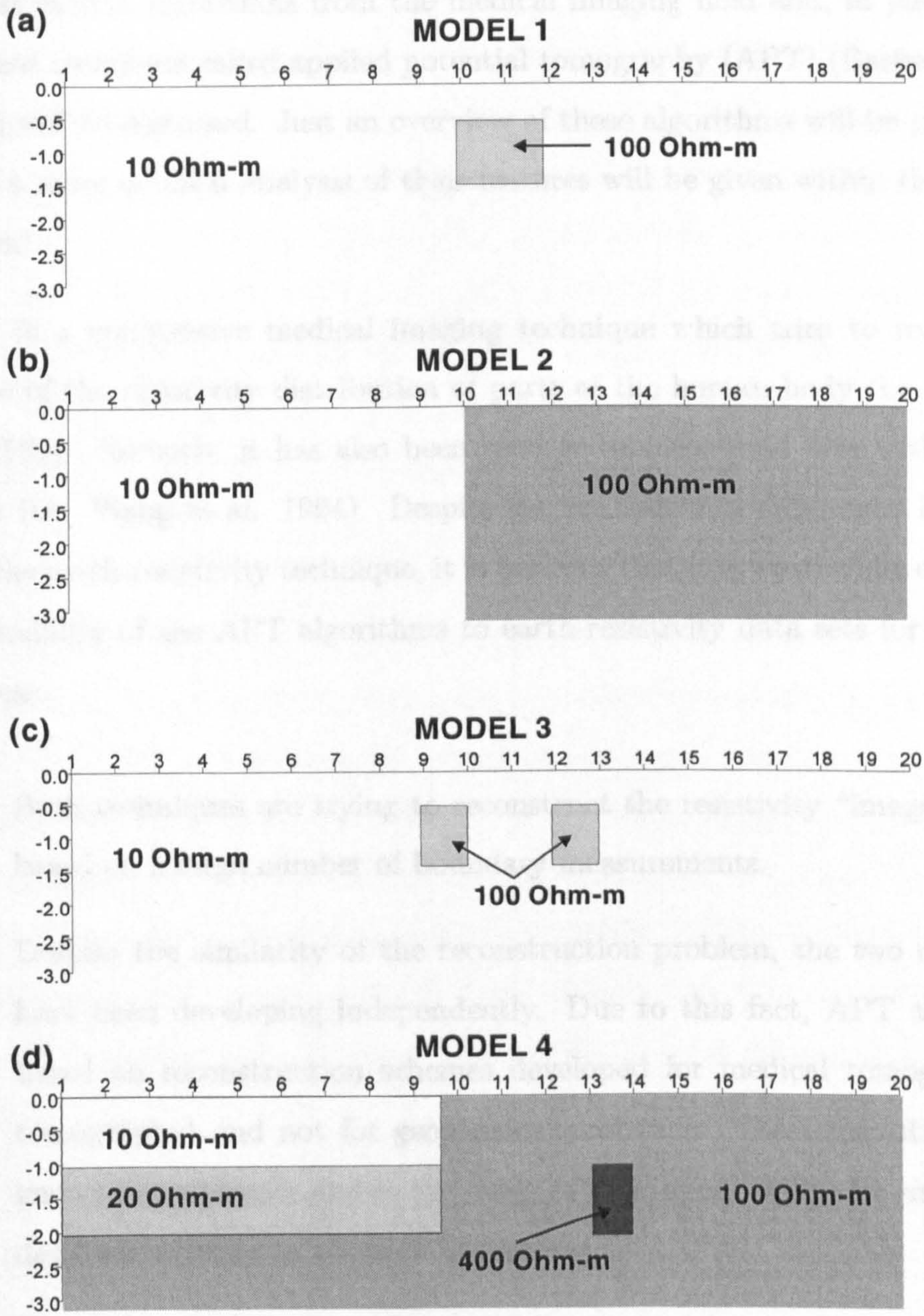


Figure 5.1: The synthetic models used in this section: a) Model 1 (one resistive prism), b) Model 2 (vertical discontinuity), c) Model 3 (two resistive prisms), d) Model 3 (conductive layers and resistive prism).

5.1.1 Inversion Algorithms from the Medical Field

In this section algorithms from the medical imaging field and, in particular, from a recent technique called applied potential tomography (APT) (Barber and Brown, 1984) will be discussed. Just an overview of these algorithms will be presented here since a more detailed analysis of their features will be given within the geophysical context.

APT is a noninvasive medical imaging technique which tries to reconstruct the image of the resistivity distribution of parts of the human body (i.e. Eyuboglu et al. 1989). Recently, it has also been used to monitor fluid flow within industrial pipes (i.e. Wang et al. 1994). Despite the fundamental differences between APT and the earth-resistivity technique, it is believed that it is worthwhile examining the applicability of the APT algorithms to earth-resistivity data sets for the following reasons:

- Both techniques are trying to reconstruct the resistivity "image" of a region based on a large number of boundary measurements.

- Despite the similarity of the reconstruction problem, the two research fields have been developing independently. Due to this fact, APT algorithms are based on reconstruction schemes developed for medical tomography (x-ray tomography) and not for geophysical problems. These algorithms are little known in geophysics and so the study of their applicability for earth resistivity data sets is likely to be quite interesting.

One of the methods widely used in APT is the back-projection (BP) method (Gordon et al., 1985), which originates within the field of x-ray tomography. In x-ray tomography the area of interest is encircled by a series of transmitters and receivers: the transmitters emit rays which are attenuated while travelling through the scanned area and are subsequently recorded by the receivers (see Figure 5.2). The recon-

struction problem can be expressed as a system of linear equations:

$$\mathbf{L}\,\mathbf{B} \;=\; \mathbf{M} \tag{5.1}$$

where **L** is the matrix of the unknown attenuation coefficients, **B** is the matrix of the weighting factors (which depend on the length of the intersection between each ray and each pixel of the scanned area), and **M** is the matrix which contains the measured ray intensities. The solution of the system by direct inversion is time-consuming since the **B** matrix is usually extremely large and ill-conditioned. One of the alternative strategies is the use of the back-projection (BP) technique. In the BP (or Summation) method the attenuation at each pixel p is obtained by summing the weighted measurements that correspond to the t rays that intersect that pixel

$$L_p \;=\; \frac{\sum_t M_{tp}\, B_{tp}}{\sum_t B_{tp}} \tag{5.2}$$

The method is the crudest x-ray reconstruction method (Herman, 1980) and, most of the time, gives results which contain major artifacts. However, the iterative versions of BP have been mathematically proven to converge to the correct solution of the linear system of equation 5.1 (Herman et al., 1973). These iterative methods are the algebraic reconstruction technique (ART) (Gordon, 1974) and the simultaneous iterative reconstruction technique (SIRT) (Gilbert, 1972). In both techniques the estimate of **L** is being corrected at every iteration k. The correction is obtained by back-projecting the difference between the original **M** and the predicted $\mathbf{M}^k$ measurements.

The only difference between the two methods is that in SIRT the corrections are made after the entire measurement set has been back-projected while in ART the corrections are made as soon as the data from only one transmitter has been back-projected (in other words, in APT, the model is updated as soon as possible). ART and SIRT have been also widely used for the inversion of borehole to borehole electromagnetic tomography data (Dines and Lytle, 1979; Radcliff and Balanis, 1979).

These methods are not limited to problems which involve only straight line paths

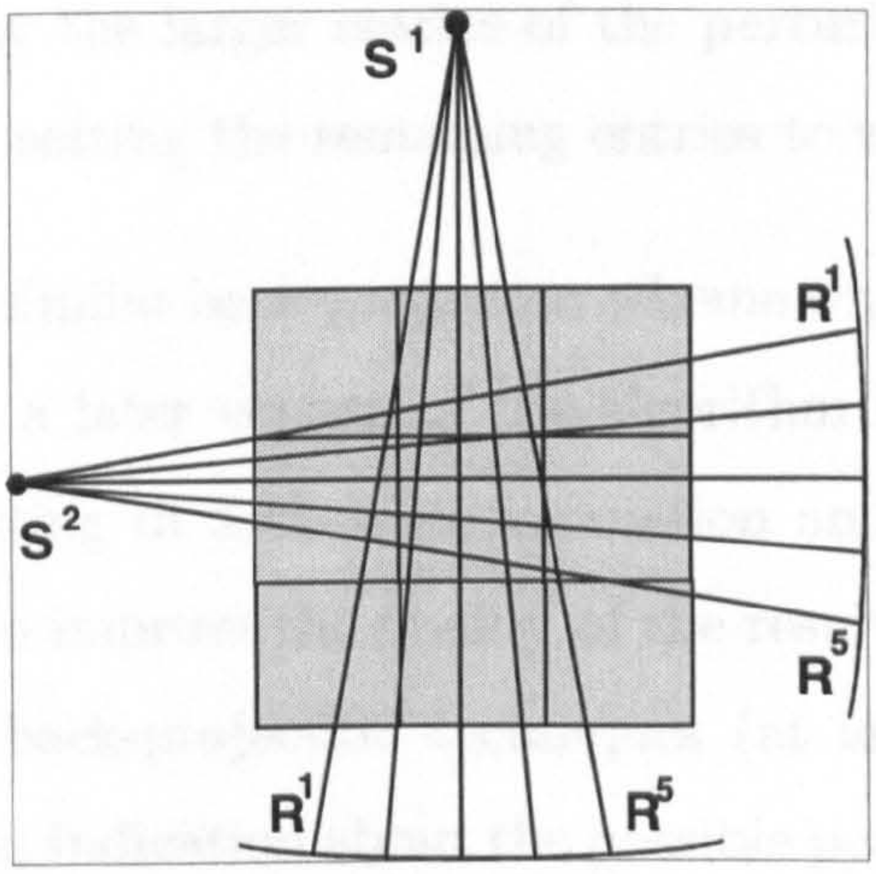


Figure 5.2: The measuring pattern in x-ray tomography. S^1, S^2 are the transmitters and $R^1, ..., R^5$ are the receivers (after Tarantola, 1987).

and may be used in other multi-transmitter/receiver reconstruction problems. This is why they were also used in the APT reconstruction problem. Barber et al. (1983) first proposed a one-step algorithm based on BP between the equipotential curves. The first version was a simple (unweighed) BP, while in later versions the weighted logarithmic ratio between the measured and calculated values was back-projected (Barber and Brown, 1987). Ider et al. (1990) suggested an iterative SIRT-type algorithm restricting the reconstruction between the equipotential curves.

Kim et al. (1983) presented an iterative ART-type algorithm which limited the BP between the current paths which end at the measuring probes. The weighting factors he used were calculated by the perturbation technique. In a later work (Kim and Woo, 1987) he suggested a weighting factor which varies linearly with the distance from the current probe. Yorkey et al. (1987a,b) modified Kim's scheme by using a SIRT-type algorithm. Furthermore he did not constraint the reconstruction between the current paths.

Tarassenko and Rolfe (1984) suggested a back-projection algorithm which made use of the perturbation matrix. This one-iteration scheme back-projects the weighted

ratio of the measured to the predicted voltage measurements. The reconstruction is constrained by using only the larger entries of the perturbation matrix (which act as the weighting factor), setting the remaining entries to zero.

Korte (1989) proposed a similar back-projection scheme which made use of the entire perturbation matrix. In a later version of the algorithm Korte (1993) showed the importance of the weighting in such a reconstruction and suggested that filtering techniques can be used to improve the quality of the resultant image. As suggested by Artola (1994) most back-projection techniques (at least those which are not iterative) can only give an indication about the possible position of inhomogeneities, and not accurate property values.

5.1.2 The Pseudosection technique

The pseudosection technique is not only the most common way to present a set of combined sounding profiling resistivity data but also the most common technique for interpreting that data - despite known deficiencies in such a simplistic approach. As explained in section 2.3.3, a sectional image of the earth is produced by projecting the apparent resistivity measurements to a point below the centre of the corresponding electrode array at a depth which is related to the probe separation. It is obvious that three main assumptions are being made:

1. The measured apparent resistivities are received as if in direct correspondence to the subsurface resistivities. Although it is obvious that in this way the correct resistivity values cannot be recovered, it is hoped that the apparent resistivity values will produce an image which will give some degree of information of about the qualitative distribution of the resistivity.

2. As far as the lateral resistivity variation is concerned, each individual measurement is assumed to give information only about the central point of the a region below the corresponding electrode arrangement.

3. As far as the vertical resistivity variation is concerned, each individual measurement is assumed to correspond to a point which has a depth that is defined by the intersection of two 45° lines which begin from two points of the electrode arrangements. These points are the centres of the current (A-B) and potential (M-N) dipoles for the dipole-dipole array, the points at the mid-distance between A,M and N,B electrodes for the Wenner array, the points of the A and M electrodes for the pole-pole array and the point of the current electrode A and the centre of the potential dipole (M-N) for the dipole-dipole array.

It is clear that the basic assumptions of the pseudosection technique are quite crude. First of all, the assumption that there is a direct link between the earth resistivity and the apparent resistivity neglects the fact (discussed in Section 4.4) that every array has areas of negative sensitivity. In a pseudosection all apparent resistivity high/low will show up as highs/lows subsurface resistivity, but it is known that due to (negative sensitivity) a low apparent resistivity measurement can be due to a resistive body and vice-versa. On the other hand, for most arrays the low sensitivity area is quite restricted so, however crude this assumption is, in a broad sense valid for a wide range of resistivity distributions and probe arrangements.

The second assumption, that of the lateral positioning of the anomaly below the centre of the array, is rather approximate, since it has been established that a fairly larger area (which also extends either side of the array) affects a measurement significantly. This assumption can still produce a reasonable approximation of the lateral distribution of resistivity for those arrays which exhibit high positive sensitivity directly under the centre of the array- namely Wenner and pole-pole. For the dipole-dipole and pole-dipole arrays this assumption is less true since they exhibit their highest sensitivity just under the potential and current electrodes. For large n separations (the current and potential electrodes are far apart) the representation of the measurement as below the centre of the array can be quite misleading (examples will be shown later).

Finally, the depth positioning of the anomaly is an arbitrary assumption since it lacks any detailed physical reasoning. This is why pseudosections are usually presented as having a depth scale which is a function of the n separation and not an absolute depth.

In Figure 5.3 an example of the pseudosections produced for a resistive prism (Model 1) is shown for different resistivity arrays. The data are presented in a grey scale contour image and the exact location of the prism is outlined within each pseudosection. Note that the traditional pseudosection depth scale is used. The results indicate the described weaknesses of pseudosection. Low apparent resistivity anomalies show up as low resistivity areas even though no such areas exist. Considering more closely the lateral and vertical positioning of the prism within the pseudosections the image for the Wenner and pole-pole arrays is quite satisfactory, but this is not the case for the remaining two arrays.

The dipole-dipole pseudosection produces a hyperbola-like high resistivity area which is clearly quite misleading. It is worthwhile explaining in detail why this dipole-dipole pseudosection image is produced. It has been established that the array exhibits a high sensitivity area in the region below the current and potential dipoles. When the potential dipole [1] is over the resistive body a high measurement is recorded. This measurement will however, be attributed to a location below the centre of the array as the distance between the centre of the array and the dipole is increased (as the n separation is increased) the anomaly will be increasingly misplaced. This generates the left branch of the hyperbola-like anomaly. Similarly the right branch of the hyperbola is created when the current dipole is over the body. It is quite clear that in the case of even slightly complicated subsurface resistivity distribution, the dipole-dipole pseudosection can be complicated and quite misleading. Such an example (Model 3-two resistive prisms) is shown in Figure 5.4a.

The pseudosection for the pole-dipole anomaly can be explained by taking into

[1]The measuring direction is from left to right and the potential dipole is placed on the right of the current dipole.

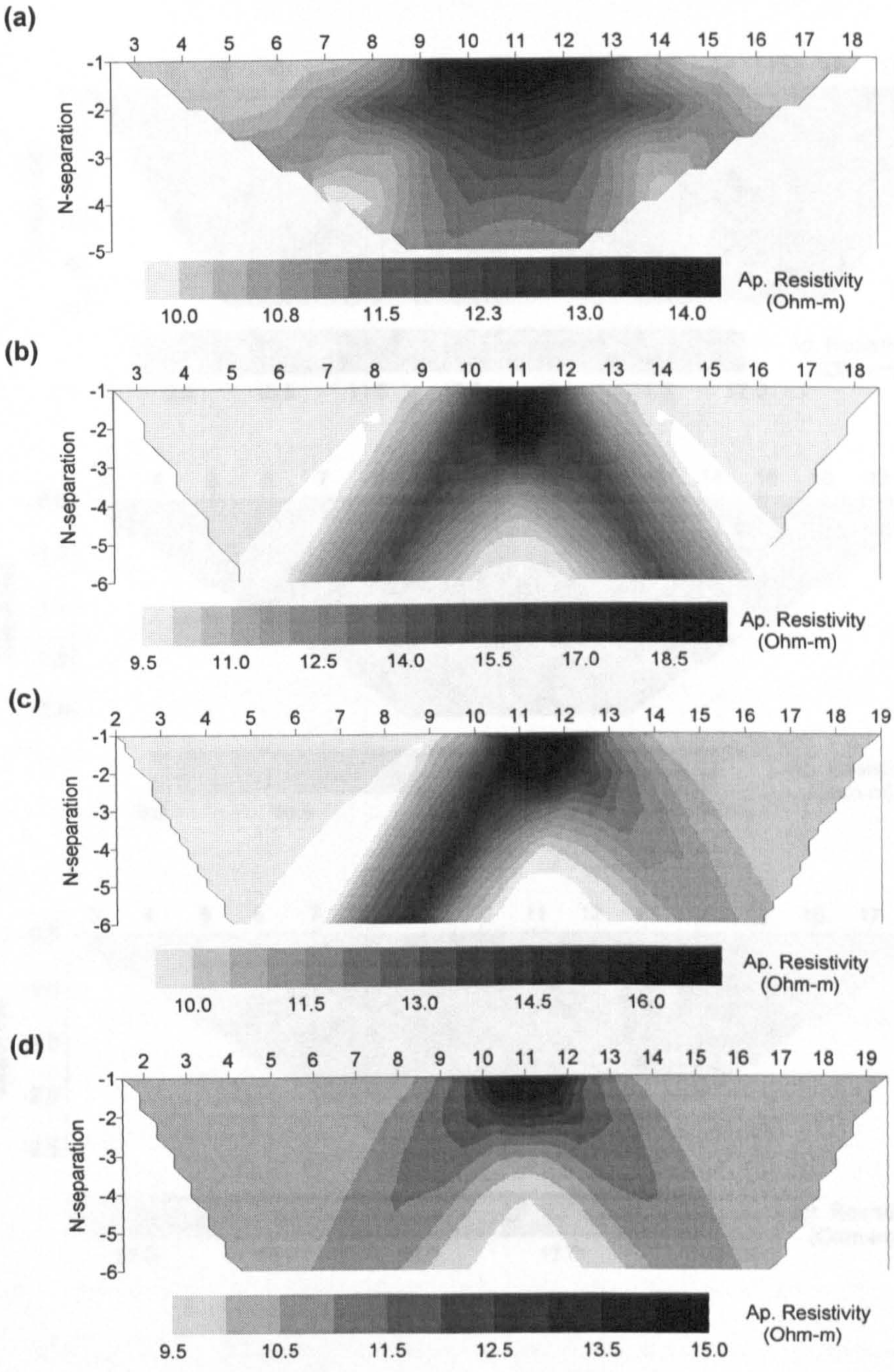


Figure 5.3: Pseudosection results corresponding to Model 1 for various arrays: a) Wenner, b) dipole-dipole, c) pole-dipole, d) pole-pole.

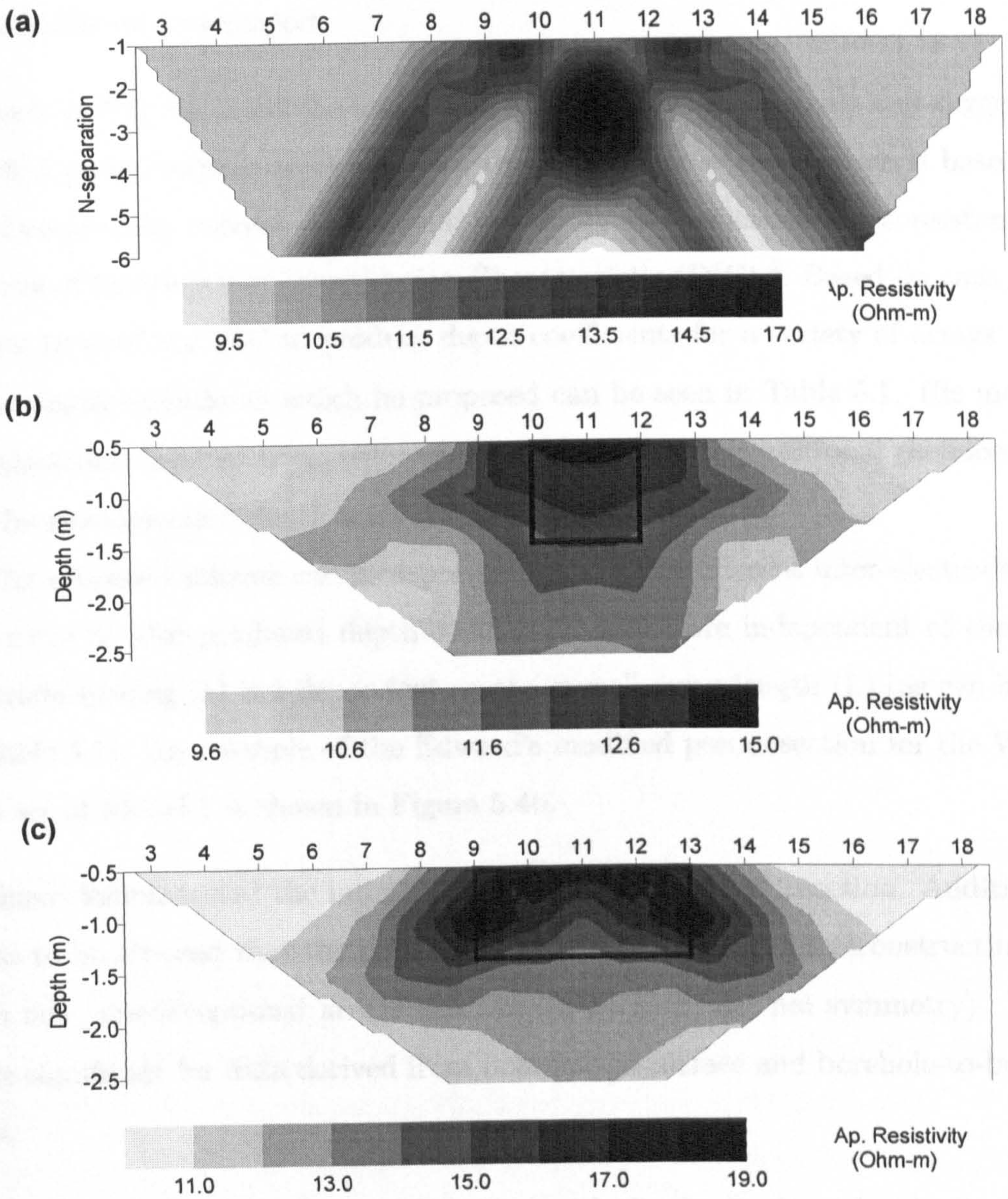


Figure 5.4: a) Dipole-dipole pseudosection for Model 3, b) Wenner modified pseudosection for Model 1, c) Wenner pseudosection for an elongated resistive prism.

account the particular sensitivity pattern of the array: a generally asymmetrical sensitivity distribution with quite a high sensitivity below the potential dipole and the current probe. The asymmetrical pseudosection image of the pole-dipole array is quite difficult to interpret.

Edwards (1977) discussed the unrealistic pseudosection depth scale and suggested a modified pseudosection depth scale for the dipole-dipole array which is based on a set of empirically derived depth coefficients. These coefficients are consistent with the idea of the Depth of Investigation Characteristic (DIC) [2]. Based on that observation, he used the DIC to produce depth coefficients for a variety of arrays. Some of the depth coefficients which he proposed can be seen in Table 5.1. His modified pseudosection has two advantages when compared to the traditional pseudosection:

1. The pseudosection depth scale is more realistic.

2. The proposed scheme can incorporate arrays with different inter-electrode spacings since he also produced depth coefficients which are independent of the inter-electrode spacing (a) but dependent on the overall array length (L) (as can be seen in Table 5.1). An example of the Edward's modified pseudosection for the Wenner data set of Model 1 is shown in Figure 5.4b.

We have demonstrated the inherent limitations of the pseudosection. Additionally, it has to be stressed that the pseudosection cannot be used for reconstructing data from new, unconventional arrays (i.e. arrays without internal symmetry). This is quite significant for data derived from borehole-to-surface and borehole-to-borehole data.

It is clear that the use of a pseudosection with dipole-dipole and pole-dipole data sets produces quite misleading results. Its use with Wenner and pole-pole data is safer but there are still situations for which the results are misleading. Such an example for the Wenner array and for the case of an elongated prism can be seen in Figure 5.4c - the pseudosection gives the impression of two resistive prisms.

---

[2]The concept of DIC and the results suggested by Edwards (1977) are fully explained in section 2.2.3.

n separation	Wenner		Dipole-dipole		Pole-dipole		Pole-pole	
	d/a	d/L	d/a	d/L	d/a	d/L	d/a	d/L
1	0.519	0.173	0.416	0.139	0.519	-	0.867	-
2	0.519	0.173	0.697	0.174	0.915	-	0.867	-
3	0.519	0.173	0.962	0.192	1.293	-	0.867	-
4	0.519	0.173	1.220	0.203	1.658	-	0.867	-
5	0.519	0.173	1.467	0.211	2.013	-	0.867	-
6	0.519	0.173	1.730	0.216	2.478	-	0.867	-
d=depth, a=inter-electrode spacing, L=external probe distance								

Table 5.1: Depth of investigation of resistivity arrays as a function of the n separation (after Edwards, 1977).

Because of its simplicity the pseudosection is widely used for initial data interpretation and in many cases is the only tool used for interpreting the data. The user however must be quite careful/experienced in order to avoid erroneous interpretations.

5.1.3 User - forward modelling interactive technique

This technique is effectively a manual trial and error procedure: the operator tries different sets of parameters and resistivity distributions within the forward modelling framework until a structure which will produce an acceptable fit between the observed and calculated data is found. It has been used extensively (Wright et al., 1985; Stretenovic and Marcetic, 1992) at a period when limited computing power did not allow the routine usage of fully automated interpretation schemes. The technique has several limitations:

1. The method is based heavily on the operator's expertise. The operator has to guess the possible structures from the original data and then try to verify these guesses. For complicated structures it is quite unlikely that this will be

successful. Additionally, the technique is subject to biased implementation.

2. Since the technique is manual there are a limited number of parameter/property combinations that the operator can try. Hence, it is quite difficult to obtain small RMS errors and there is no guarantee that an optimum fit is achieved.

3. Most importantly, the technique cannot take advantage of high resolution data sets (such as these produced by new automated measuring systems) since it is impossible to handle manually a large number of parameters. Therefore, only a rough/crude idea of the subsurface resistivity distribution is obtained.

On the other hand, the technique is quite easy to implement and, if used properly, it can produce a useful initial/general subsurface resistivity image. In this way it can help to create an acceptable parameter set for subsequent use with fully automated interpretation schemes.

5.1.4 The Zhody-Barker technique

Barker (1992) modified an algorithm suggested by Zhody (1989) for 1-D resistivity data and developed an scheme which is effectively an iterative pseudosection technique.

Barker used the finite difference forward modelling scheme proposed by Dey and Morrison (1979a) as a base of his iterative technique. He generated a set of parameters consistent with the pseudosection by considering each pseudosection point as the centre of a parameter region. He used the DIC concept to define the depth of each pseudosection point (and thus the depth of each parameter centre). The limits of each parameter were then found by linear interpolation. Note that in this way there is a one-to-one correspondence between a parameter region and a measurement. By assuming that each measurement is directly affected by only its corresponding parameter the iterative procedure is as follows:

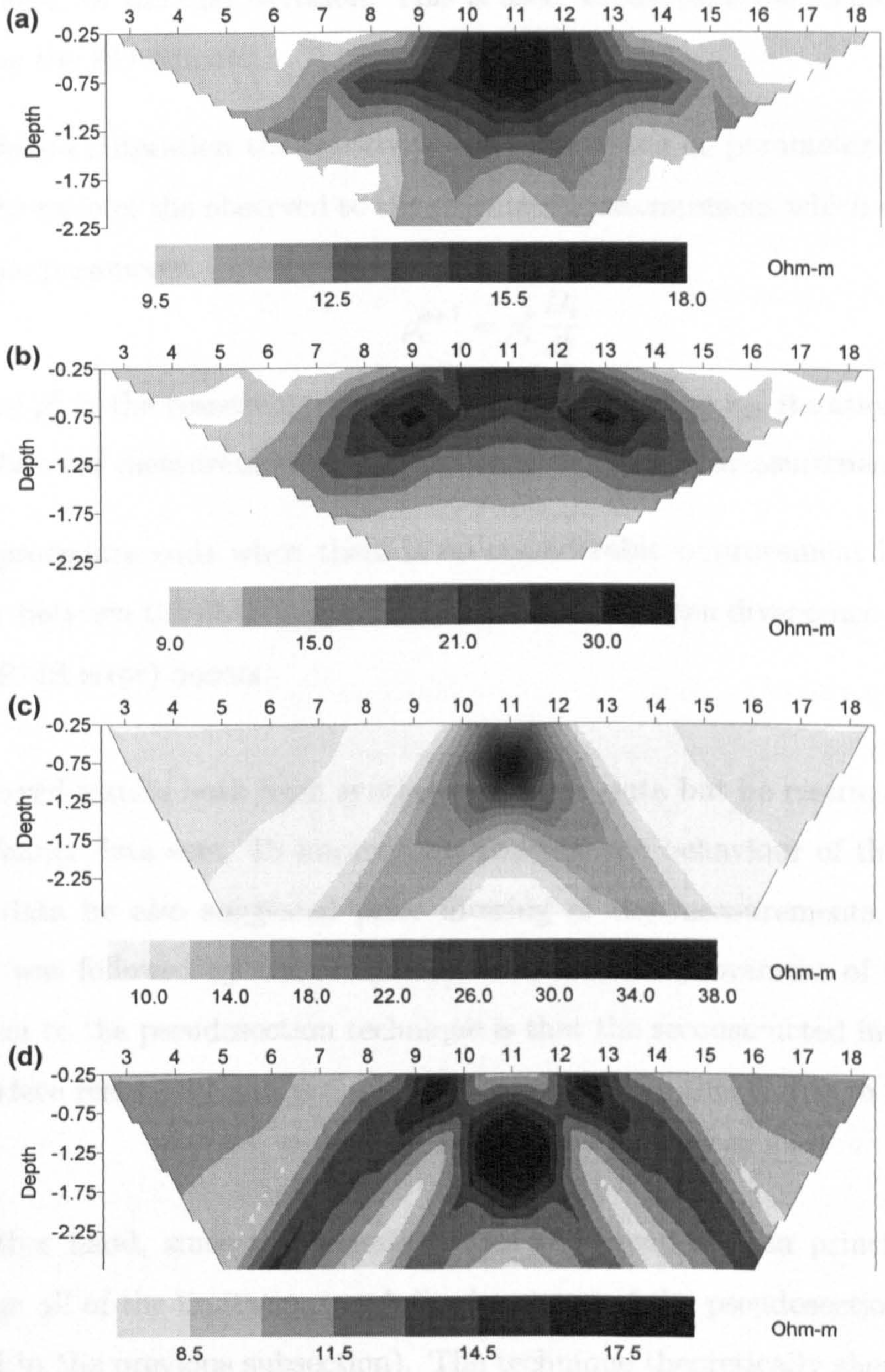


Figure 5.5: Barker's method: a) reconstruction of Model 1 (Wenner array), b) reconstruction of an elongated resistive prism (Wenner array), c) reconstruction of Model 1 (dipole-dipole array), d) reconstruction of Model 3 (dipole-dipole array).[All calculations were performed by the author.]

- The initial pseudosection image is used to produce a starting resistivity distribution for the first iteration. This is used to calculate the forward response (using the FD scheme)

- At the k+1 iteration the resistivity estimate of the i_{th} parameter is corrected by the ratio of the observed to the calculated measurement which corresponds to this parameter.

$$\rho_i^{k+1} = \rho_i^k \frac{D_i}{d_i^k} \tag{5.3}$$

  where ρ_i^k is the resistivity of the i_{th} parameter for the k_{th} iteration, D_i is the i_{th} observed measurement and d_i^k is the i_{th} calculated measurement.

- the procedure ends when there is no considerable improvement in the RMS error between the observed and modelled data or when divergence (increase of the RMS error) occurs.

Barker showed results both from synthetic and real data but he restricted his work only to Wenner data sets. To improve the convergence behaviour of the algorithm for noisy data he also suggested prior filtering of the measurements. A similar procedure was followed by Dahlin (1993). The main improvement of the method with respect to the pseudosection technique is that the reconstructed image depicts the subsurface resistivity and not apparent resistivities - this is due to its iterative nature.

On the other hand, since the technique uses the pseudosection principles it still suffers from all of the limitations and disadvantages of the pseudosection technique (described in the previous subsection). The technique theoretically should perform satisfactorily in cases where the pseudosection works, but will produce artifacts in cases where the pseudosection does not work.

We performed several reconstructions using the Zhody-Barker method in order to verify this in practice. The FEM forward model was used as the basis of the reconstruction algorithm and we followed the parametrization suggested by Barker (1992).

The results are depicted in Figure 5.5. For the case of Model 1 the pseudosection image for the Wenner array (Figure 5.4b) is quite satisfactory (good delineation of the borders of the prism). In this case the results of Barker's method, shown in Figure 5.5a are satisfactory as well. The method took 6 iterations and produced an RMS error of 3%. But, for the case of the elongated prism (Figure 5.4c) Barker's method, as was expected, did not perform satisfactorily (see Figure 5.5b) - the reconstruction of the Wenner data produces two large anomalies which give the wrong impression of two prisms. In this case the algorithm diverged after 5 iterations.

Moreover, by following an equivalent parametrization scheme for the dipole-dipole array we were able to test Barker's method for other arrays apart from Wenner. The results for the reconstruction of the Model 1 for the dipole-dipole data set are shown in Figure 5.5c and it is clear that the reconstruction does suffer from similar artifacts to those produced by the pseudosection approach. Disappointing results were obtained from the reconstruction of the dipole-dipole data for Model 3 (two resistive prisms) depicted in Figure 5.5d. In this case the algorithm diverged after 2 iterations.

The tests indicated that the technique has no self-correcting mechanisms as far as the location of the target distribution is concerned. Further, the technique cannot cope with unconventional arrays. An additional problem with Barker's method is that the method cannot cope with data sets that have missing data points. If there is a missing data point [3] since there is a one-to-one mapping between a measurement and a parameter, there will be a parameter which cannot have a resistivity value. But the forward modelling scheme requires values for all parameters in order to operate.

---

[3]This is quite common when surveying. Due to natural or man made obstacles there might be positions where a probe cannot be inserted this will result in missing many measurements. For example a Wenner profile for 20 probes (n=1) results in 17 measurements. If an intermediate probe cannot be inserted (19 probes in total) then only 13 measurements can be taken.

5.1.5 The Bristow technique

Another approximate reconstruction technique was proposed by Bristow (1966). The method is used only for pole-dipole data and has been used extensively used for locating cavities (Smith, 1986). The pole-dipole measuring scheme used is somewhat different from the typical combined sounding profiling procedure since it involves all the possible pole-dipole measurements that can be measured (see Figure 5.6a).

In this method, measurements with identical probe spacings form a profile within which regions of anomalous resistivity are located. Using the current electrode positions as centres, arcs enclosing the resistivity anomalies are constructed. The intersections of the arcs define the source of the anomaly. These arcs are actually the equipotential lines (for homogeneous ground) that end on the edges of the area of the high anomaly (see Figure 5.6b). An example of how the reconstruction operates is shown in Figure 5.6c.

However, the method is quite crude for the following reasons: the target location procedure is conducted graphically and therefore no indication of the amplitude of the anomaly can be incorporated. In this way a significant amount of information is lost and, furthermore, misinterpretations may occur. Moreover the graphical location of the anomalies renders the method subject to inaccuracies. That is why Bristow's method was mostly used for locating resistivity anomalies rather than for imaging the electrical properties of the subsurface. On the other hand, the technique is quite simple and has produced successful results in many cases and recent studies (Lowry and Shive, 1990) claim that the technique can be used successfully to delineate the shape of the target as well.

As suggested by Lowry and Shive (1990), The technique has been criticised for being based on faulty assumptions and at first sight this seems to be reasonable since the introduction of the arcs/equipotential lines enclosing the high anomalous area is quite arbitrary and lacks physical reasoning. However this is not the case: the technique has a physical reasoning which is not directly related to the arc/equipotential

lines but to the sensitivity of the points that lie between the arcs/equipotential lines. In Figure 5.6b it can be seen that for a pole-dipole measurement the equipotential lines enclose the area for which the sensitivity is quite high.

If a high measurement is observed for this arrangement then this should be due to a feature within the high sensitivity region (enclosed by the arcs) and thus the anomaly should be (roughly) located there; consequently the graphical summation of these "anomalous regions" will tend to reveal the surveyed feature. In other words, the technique works for the pole-dipole array simply because a great part of the high sensitivity region of the array is coincidentally located between the equipotential lines. It is a coincidence since there is no physical reason for the equipotentials to enclose the high sensitivity region.

As will be shown in the next section Bristow's method is nothing but a simple graphical back-projection. The fact that the region between the equipotential lines also includes an area which is less sensitive (see Figure 5.6b) suggest that if the technique is used for full imaging it will produce undesirable artifacts.

5.1.6 Back-projection techniques

In this part the back-projection techniques proposed for the approximate reconstruction of surface resistivity data set will be reviewed. The physical meaning of the back-projection was mentioned when discussing the Bristow method: the basic idea is to project each individual measurement to an area which exhibits the maximum sensitivity. For every measurement a subsurface "image" is created. The final reconstruction "image" is the summation of these elementary "images".

The back-projection technique for surface linear arrays was first used by Powell et al. (1987) who modified the algorithm of Barber et al. (1983). The technique uses a measuring technique similar to the dipole-dipole continuous sounding profiling scheme. The reconstruction is a simple BP constrained between the area of the

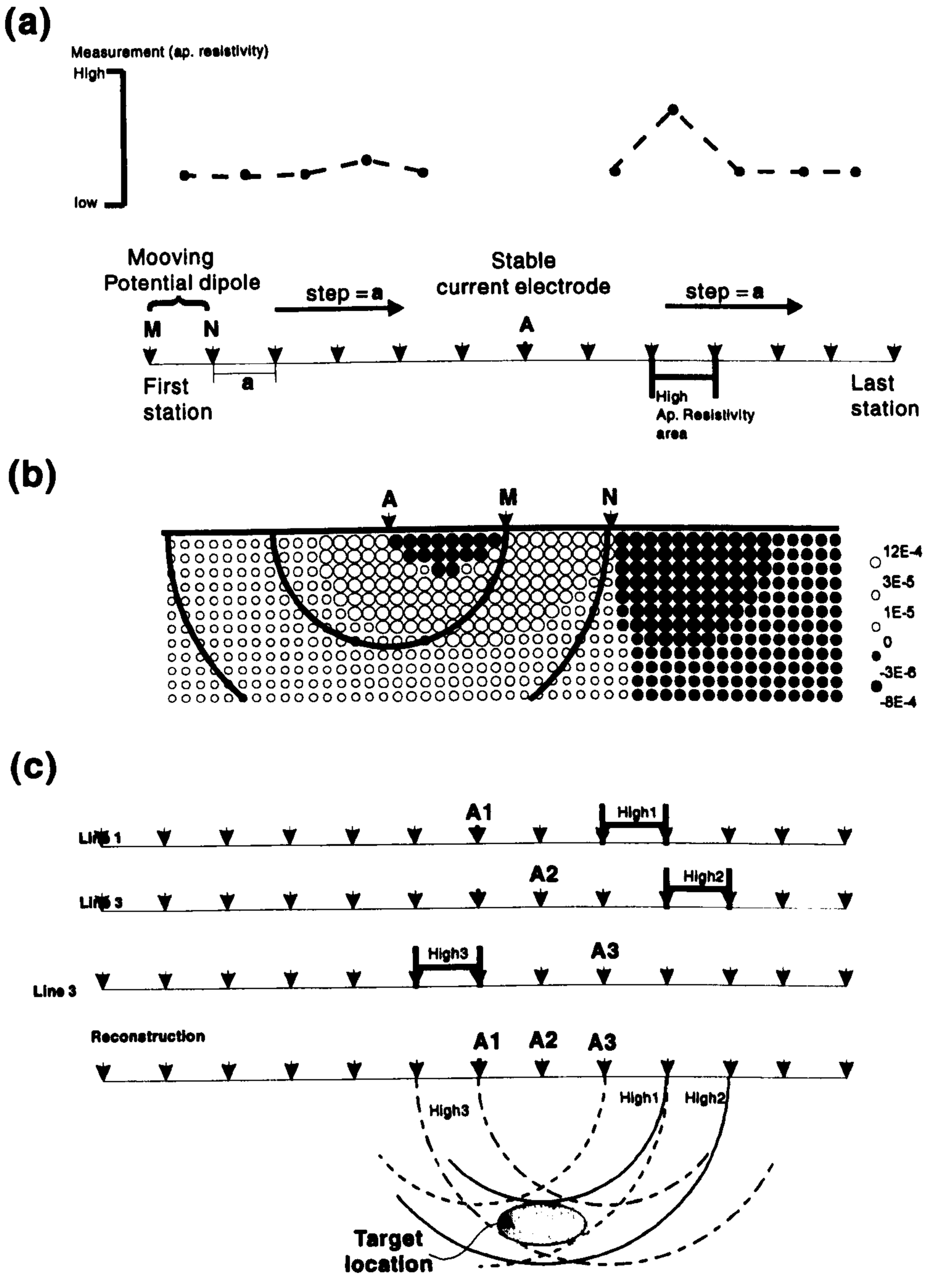


Figure 5.6: Bristow's technique: a) the measuring scheme, b) the equipotential lines (arcs) in relation to the sensitivity matrix, c) an example of how the method works.

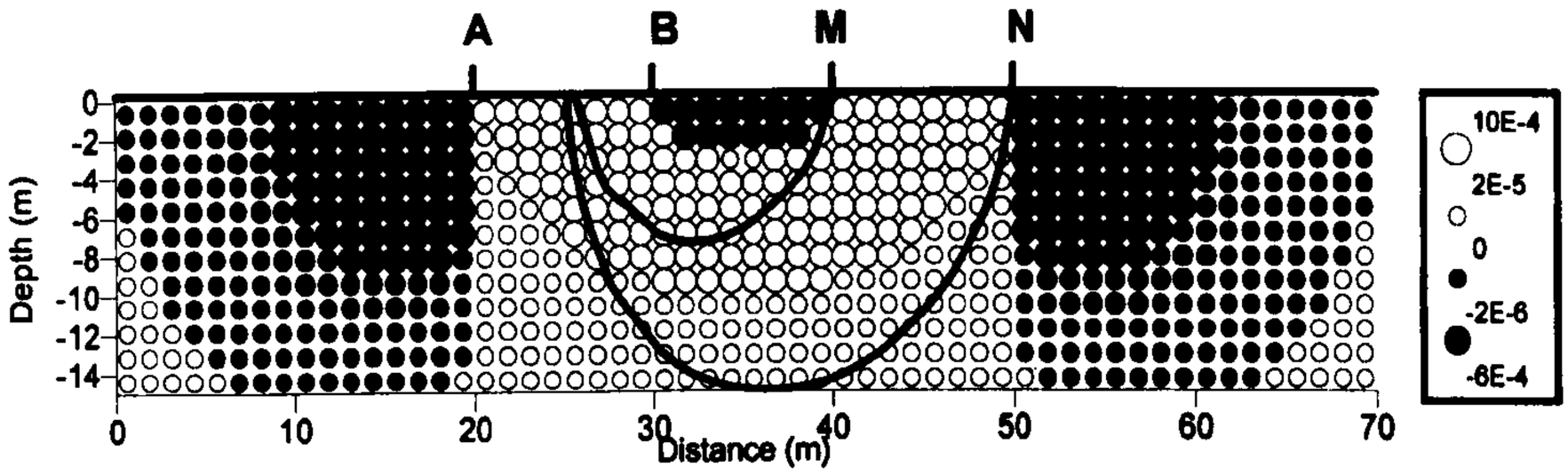


Figure 5.7: Equipotential lines and associated sensitivity for the dipole-dipole array.

equipotential curves which end at the potential dipole. The resistivity ρ_i of each block i is given by the following formula:

$$\rho_j = \rho_j^o \sum_i \frac{V_i^m}{V_i^o} \delta_{ij} \tag{5.4}$$

where V_i^o is the i_{th} measurement for homogeneous ground of resistivity ρ_j^o, V_i^m is the observed i_{th} measurement, and δ_{ij} is 1 if the centre of the j_{th} parameter is between the equipotential lines which end at the potential electrodes of the i_{th} measurement otherwise is 0.

Powell et al. (1987) presented successful model reconstructions which indicate the potential of the method for linear earth-resistivity arrays. Although the method is fully computerized it is quite clear that is not very much different to the Bristow's method. The area between the equipotential lines (calculated for a homogeneous earth) that end at the potential electrodes, for the dipole-dipole array is the area of the highest sensitivity, as can be seen in Figure 5.7.

Noel and Xu (1991) and Noel (1992) presented a modified version of the BP algorithm by introducing a weighting factor which is actually an estimate of the sensitivity matrix calculated analytically. The measuring scheme that Noel followed is not a conventional one since it involves all of the possible independent measurements that can be obtained when a series of probes is used (Xu and Noel, 1991). However, this scheme does not, in fact, turn out to be much different from the classical dipole-dipole measuring scheme. The one-step algorithm assumes an initial homogeneous resistivity distribution and back-projects the weighted ratio of the measured to cal-

culated (for homogeneous earth) potential to the area between the equipotential curves using a formula very similar to equation 5.4.

$$\rho_j = \rho_j^o \frac{\sum_i (V_i^m / V_i^o)\, W_{ij}}{\sum W_{ij}} \delta_{ij} \tag{5.5}$$

the extra term W_{ij} is the Jacobian matrix entry for the i_{th} measurement and j_{th} block. Weighting the back-projection using the Jacobian matrix results in taking into account the quantitative contribution of each parameter to the measurement. On the other hand, the weighting implies a direct linear relationship between the measurement and the Jacobian matrix, which is not true.

Since both the homogeneous potentials and the sensitivity matrix are calculated using the analytical solution of Laplace's equation (see equation 2.30) the algorithm cannot be iterative.

Noel presented several results for synthetic and real data. He observed that at the side edges of the subsurface image the quality of the reconstruction is quite poor and he suggested reconstructing only the central region of the area of interest. The reconstructions are better than the respective pseudosections but still suffer from artifacts (Tsourlos, 1992).

We applied the technique to a series of models but we faced a problem when we tried to reconstruct dipole-dipole data sets: the region between the equipotential lines is asymmetrical (it is larger in the region near the potential probes) but the sensitivity for the dipole-dipole is symmetrical (see Figure 4.4c). This resulted in an asymmetrical reconstructed image. This problem was tackled by also including into the reconstruction the reciprocal measurement (see Section 2.1.6) for each data-point (which is, by definition, identical to the original one).

In Figure 5.8a,b reconstruction examples of dipole-dipole data sets for a resistive prism and a vertical discontinuity can be seen. It is clear that the prisms still produce a hyperbola-like image, although not as pronounced as that of the pseudosection.

One of the disadvantages of the methods that constrain the BP between equipoten-

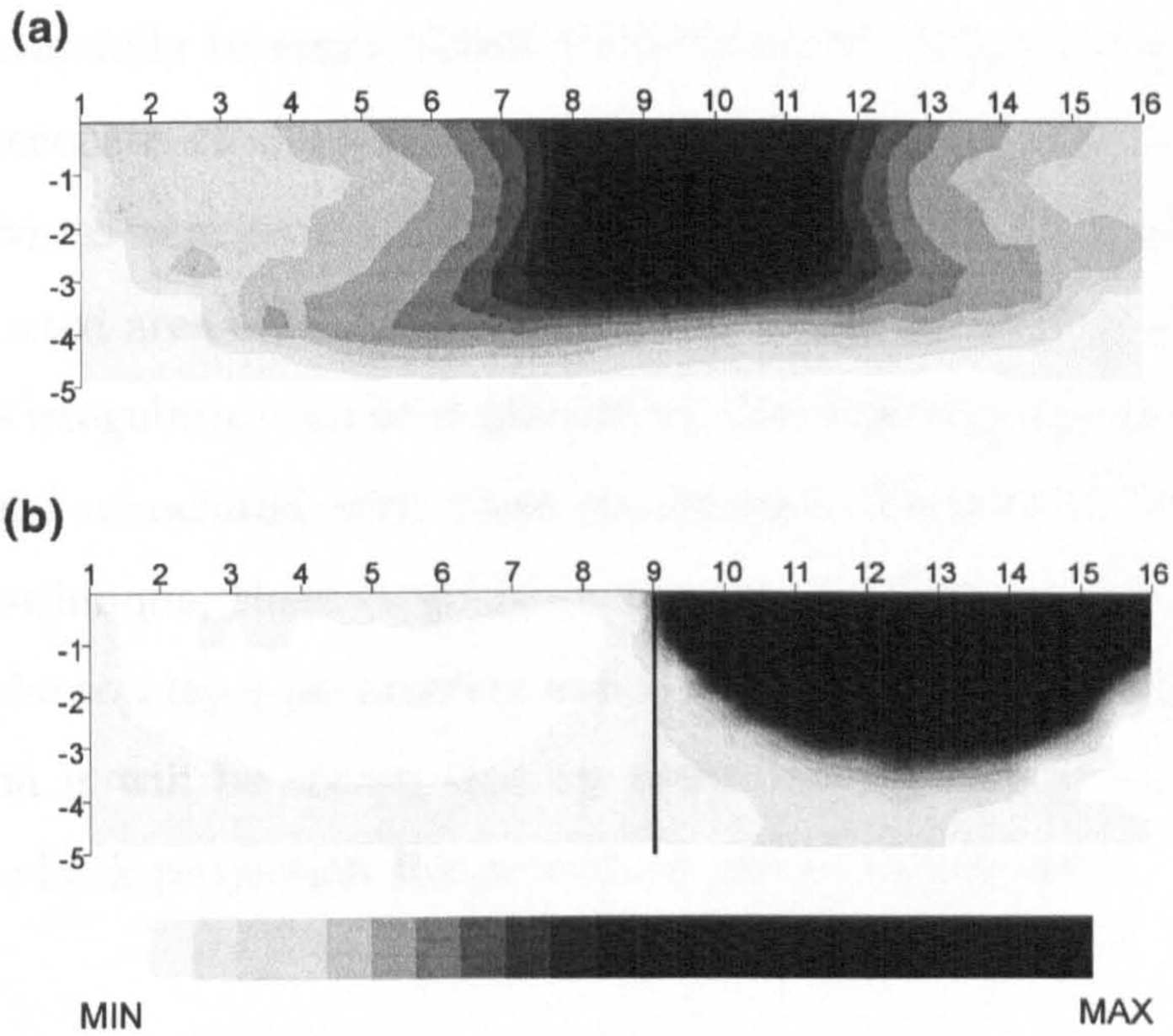


Figure 5.8: Example using Noel's method: a) reconstruction of a resistive prism (dipole-dipole data), b) reconstruction of a vertical discontinuity model (dipole-dipole data). The limits of the models are outlined within the image. The resistivity of the models is 100 Ohm-m and the background is 10 Ohm-m [The calculations were performed by the author].

tial lines (such as Noel's method) is that no data from arrays such as Wenner or pole-pole can be reconstructed. Especially for the pole-pole array, the equipotential line constraint is meaningless.

Shima (1992) used a BP algorithm to reconstruct crossborehole-to-surface data sets. He used a one-step back-projection formula which is identical to the one described by equation 5.5 but he did not apply any constraint to the BP: he used the entire Jacobian matrix (both positive and negative values). The synthetic results which he presented were quite satisfactory and no major artifacts appeared. The only problem which he mentioned was that the introduction of the negative sensitivity matrix entries into the BP caused irregularities in the resistivity distribution at the corners of the reconstruction when the size of the parameters was too small.

We tried unsuccessfully to apply Shima's algorithm to surface resistivity data sets. Although the reconstructed image was quite close to the original structure, invalid negative resistivities were produced for those parameters which are located at the top of the reconstructed area - this did not change when the size of these parameters was increased. This irregularity can be explained by the very large values of the Jacobian matrix which are associated with these parameters. Because of the approximate nature of the technique, these large Jacobian values will exaggerate the resistivity correction for the top layer parameters and will produce "unreasonable" results. In the next section it will be shown that by excluding some of the Jacobian matrix values from the back-projection the procedure can be stabilized.

5.1.7 Other techniques

Beard and Morgan (1991) suggested the use of 1-D inversion techniques applied to 2-D structures. The idea is based on the fact that every combined sounding profiling measuring scheme can be viewed as a series of continuous individual soundings (see Section 2.3.1). By performing a 1-D interpretation of each individual sounding a series of adjacent resistivity columns is obtained. Simple interpolation can then produce a "pseudo-2-D image" of the earth. Their results indicate that the technique is successful in delineating the top and bottom edges of 2-D structures but, as expected, failed to give satisfactory results for the sides of the structures. Dahlin (1993) used this technique extensively to reconstruct Wenner CSP data. He observed that the technique can be sensitive to noise, so he suggested a procedure for producing smoothed results. He concluded that the technique is extremely fast and can produce valid results for layered-like structures however the user should bear in mind that the interpretation is strictly 1-D and thus the technique was inaccurate in locating the lateral extension of 2-D targets.

Li and Oldenburg (1991) presented an approximate fast inverse algorithm based on linear programming. A Born approximation was used as the forward modelling tech-

nique and hence the suggested reconstruction reduced to a one-step matrix inversion procedure (at least for the 2-D model he presented). The presented modelled data reconstructions were particularly good. A similar algorithm has been presented by Xu (1993). The algorithm was fast and produced reasonably good results for simple structures. However, its ability is limited by the lack of resolution when the resistivity contrasts are high (Xu, 1993) which is a consequence of the Born approximation.

Finally, Hobbs and Reading (1994) presented a 2-D mapping technique which takes advantage of the sensitivity variation of the Wenner array [Offset Wenner, (Barker, 1981)] and used it to locate conductive zones associated with faults. By its nature this technique has a limited applicability to this particular type of structure.

5.1.8 Conclusions

In this section widely used techniques for the approximate reconstruction of earth-resistivity data have been reviewed. The conclusions of this study can be summarized as follows:

The pseudosection technique is based on crude assumptions but it is simple in its implementation. It produced results which, in most cases, suffer from major artifacts which are associated with the varying sensitivity of the arrays. For arrays such as Wenner and pole-pole (which have high positive sensitivities below their central region) the pseudosection working more satisfactorily. But, for more complicated structures as well as for arrays such as the dipole-dipole and pole-dipole it produces misleading results. Further, the technique cannot cope with unconventional measuring schemes (i.e. full tomographic data sets).

The associated Barker's method suffers from the same limitations. Despite its iterative character it has no self-correcting mechanism for adjusting the lateral resistivity distribution. It is able to produce reliable reconstructions (both in qualitative and quantitative fashion) only when the pseudosection assumptions are valid.

The user-forward modelling interactive technique is based heavily on the operator's expertise and is subject to biased implementation. Further, it cannot cope with the increased amount of data produced by automated measuring systems.

The simple back-projection algorithms (such as Bristow's method and Powell's method) are quite crude but easily applicable since there is no need to calculate the Jacobian matrix. The equipotential constraint however can be applied to a limited number of arrays. When it works it guarantees that BP will take place at the most sensitive regions of the subsurface. The associated Noel's method is more sophisticated (it makes use of the Jacobian matrix) but still suffers from reconstruction associated artifacts.

Overall, in this section the major importance of the Jacobian matrix within the approximate reconstruction procedure was demonstrated. The sensitivity pattern of the various probe arrangements can explain the limitations of the various techniques and most importantly can give a physical insight into why these techniques work. Based on these observations a more general approximate technique can be proposed as it will be shown in the next section.

5.2 A Generalized Iterative Back-projection Algorithm

In this section a generalized back-projection algorithm for reconstructing earth-resistivity data is presented. The need for such an algorithm derives from an analysis of the limitations of the existing approximate algorithms.

The presented algorithm is characterized as general for several reasons.

a) It can include all of the presented BP algorithms (the methods of Bristow, Powell, Noel, and Shima) as well as techniques such as the pseudosection and Barker's method.

b) It can be iterative and, in this way, can produce a quantitative "image" of the subsurface resistivity.

c) It can reconstruct data from any array as well as from unconventional probe arrangements and full tomographic data. Further, it can reconstruct incomplete data.

This is achieved by recognizing the major importance of the Jacobian matrix within the approximate reconstruction procedure. Therefore, the constraints within the reconstruction are defined in relation to the JM and not as arbitrary expressions (e.g. equipotential lines constraint).

5.2.1 The algorithm

The proposed iterative technique seeks to obtain an estimate of the subsurface resistivity distribution for which the predicted measurement values (obtained by the solution of the forward problem) are as close as possible to the measured data.

In order to achieve this, an initial resistivity distribution is assumed (usually uniform) and by using the forward modelling technique, the modelled measurements that correspond to this distribution are obtained. These modelled measurements are compared with the original data and the weighted differences are back-projected in order to obtain a correction to the resistivity estimate. This correction is added to the current resistivity distribution and the procedure is repeated until the difference between the measured and the modelled data satisfies a stopping criterion.

Let the M measured data values obtained by using a tomographic measuring scheme be represented by a vector $\mathbf{D}$ where $\mathbf{D}^T = \{D_1, D_2, ...D_M\}$, and let $\mathbf{d}$ be the vector which represents the modelled data, $\mathbf{d}^T = \{d_1, d_2, ...d_M\}$. The N subsurface blocks that are allowed to vary their resistivity independently (the unknown parameters of the problem) are represented by a vector $\mathbf{x}$ with $\mathbf{x}^T = \{x_1, x_2, ..., x_N\}$. The iterative procedure can be defined as:

1. $\mathbf{x}^k$ is the resistivity estimate at the k^{th} iteration;

2. Calculate the modelled data set $\mathbf{d}^k$ that corresponds to this resistivity distribution by using the FEM;

3. Find the new resistivity vector $\mathbf{x}^k$: the new resistivity estimate at every block j will be

$$x_j^{k+1} = \frac{\sum_{i=1}^{M} (D_i/d_i^k)\, W_{ij}^k}{\sum_{i=1}^{M} W_{ij}^k}\, x_j^k \tag{5.6}$$

where W_{ij} is the weighting factor, which for our case is the Jacobian matrix entry for the k_{th} iteration which corresponds to block j and measurement i.

4. Repeat until the RMS between the measured and modelled data is practically stable (i.e. less than 1% improvement), or divergence occurs, or if a preset number of iterations has been reached.

This procedure has similarities to the simultaneous iterative reconstruction technique (SIRT) in the sense that the resistivity corrections are taking place after the entire data set has been back-projected. Equation 5.6 describes the back-projection procedure. The final correction factor for each block is the sum of the weighted ratios between the observed and the modelled data. The updated resistivity estimate is obtained after multiplying the correction factor with the current resistivity estimate and in that sense this procedure can be described as a multiplicative back-projection. It will be shown that by choosing appropriately the weighting factors and constraints in equation 5.6 a wide variety of techniques can be described.

Alternatively, iterative additive back-projection can also be used for the resistivity reconstruction problem (e.g. Tsourlos et al. 1993) with similar results. In such a case the corrected resistivity is given by:

$$x_j^{k+1} = x_j^k + \frac{\sum_{i=1}^{M} (D_i - d_i^k)\, W_{ij}^k}{\sum_{i=1}^{M} W_{ij}^k} \tag{5.7}$$

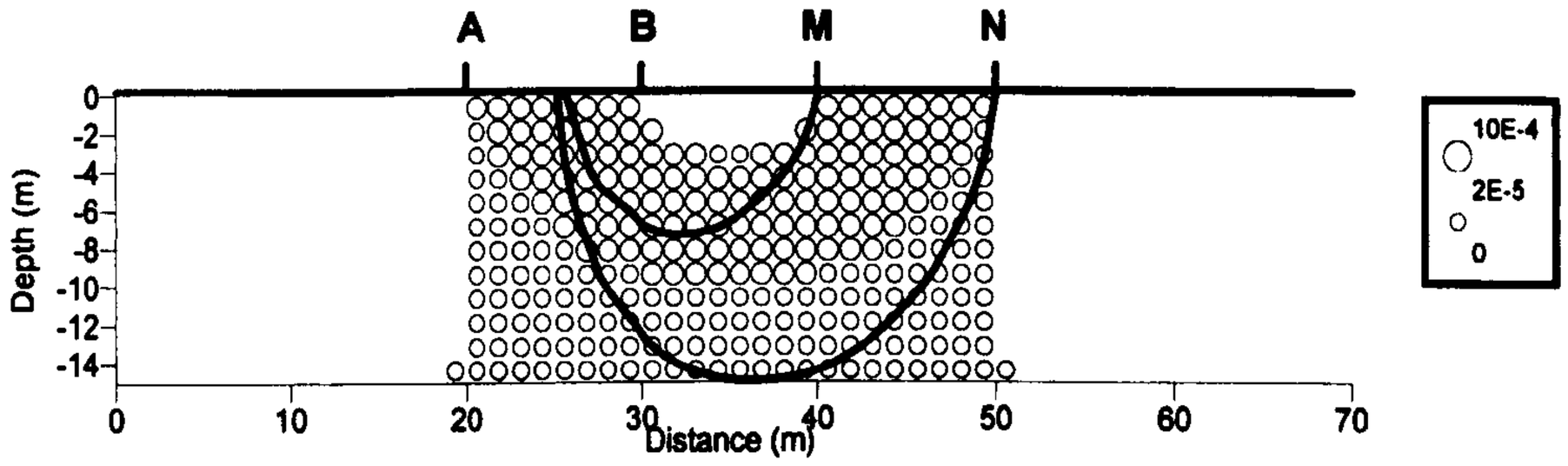


Figure 5.9: A comparison between the equipotential line and the generalized constraint.

Equation 5.6 can be very easily modified in order to become an ART-type algorithm: the only difference is that the correction takes place after only one set of measurements (the set that contains measurements which have identical current probes) is back-projected. However, tests with such an algorithm were not successful since the procedure was very sensitive to noise (Tsourlos, 1992).

The weighting factor can play a significant role in the BP procedure since each subsurface parameter contributes to each individual surface measurement to a different extent. By using parts of the JM as a weighting factor this varying sensitivity of the parameters towards the measurements is reflected to the reconstruction. In this scheme the Jacobian matrix was calculated using the adjoint equation approach. Since in every iteration the resistivity distribution changes it follows that the Jacobian matrix changes as well. Therefore, the Jacobian matrix is updated in every iteration.

5.2.2 Constraining Back-Projection

It has been explained that for linear resistivity arrays the use of the entire JM within the BP procedure can result in negative subsurface resistivities (see section 5.1.6). In order to tackle this problem it is common to apply some constraints and to use only parts of the Jacobian matrix. Instead of using arbitrary constraints such as those of the equipotential lines, we suggest the use of a constraint which is directly

related to values of the Jacobian matrix. Taking into account the direct link between the JM and the BP procedure, such a constraint allows the application of the BP for any type of resistivity data.

If a threshold value t is defined, the weighting factor W_{ij} of equation 5.6 is given by

$$W_{ij} = \begin{cases} J_{ij} & \text{if } J_{ij} \geq t \\ 0 & \text{otherwise} \end{cases} \tag{5.8}$$

This threshold value can be decided after several tests. In particular, model tests were carried out in order to find the optimum value which is the highest possible one that does not affect the produced image. For most arrays the value t=0 (which excludes all the negative sensitivity entries) was a valid choice.

In Figure 5.9 a comparison between this constraint and the equipotential line constraint for the dipole-dipole data can be seen. It is clear that this proposed threshold allow a larger part of the sensitivity matrix to participate in the reconstruction, and that the equipotential constraint is an overally high threshold.

As it was already stated, the scheme described by equation 5.6 can be used to describe a variety of techniques. In particular, one iteration of the described algorithm is equivalent to Noel's method if the weighting factors are changed. If Q is the area between the equipotential curves which end at the measuring probes the weighting factor will be

$$W_{ij} = \begin{cases} J_{ij} & \text{if the centre of the parameter } j \in Q \\ 0 & \text{otherwise} \end{cases} \tag{5.9}$$

Note that in such a case an iterative equipotential constraint procedure can also be defined (i.e. Yorkey (1986)). Similarly, one iteration of the scheme is equivalent to Powell's and Bristow methods if the following constraint is used:

$$W_{ij} = \begin{cases} 1 & \text{if the centre of the parameter } j \in Q \\ 0 & \text{otherwise} \end{cases} \tag{5.10}$$

Finally, the pseudosection procedure can be defined in similar way: if P is the

pseudosection projection point one iteration of the scheme is equivalent to the pseudosection if the following constraint is used:

$$W_{ij} = \begin{cases} 1 & \text{if the centre of the parameter } j = P \\ 0 & \text{otherwise} \end{cases} \tag{5.11}$$

The iterative scheme using that constraint is Barker's method. From this it is quite clear that the pseudosection method is effectively a crude back-projection procedure.

5.2.3 Reconstruction examples and discussion

A computer program was written in order to enable the automatic reconstruction of surface resistivity data. The FEM was used as to calculate the forward model response and the Jacobian matrix was calculated using the adjoint equation technique (section 4.3.2). The size of each parameter region was set to be a square with a side of half the inter-electrode spacing. The number of parameters in each layer is symmetrically reduced as depth increases since the parameters at the edges are not well-resolved. Outside the parametrized area the element resistivities were set to be equal to that of the nearest side or bottom parameter. Several tests were conducted to validate the algorithm.

The dipole-dipole data set for Model 1 was reconstructed using one iteration of the generalized algorithm. The results are shown in Figure 5.10a. In Figure 5.10b the results for the same model after 10 iterations are shown. The quantitative results are not that close to the original ones. Despite this fact there is no doubt that the iterative procedure delineates the target far better than the non-iterative scheme. The convergence behaviour for this reconstruction can be seen in Figure 5.11: during the first few iterations the RMS error reduces rapidly while in the following iterations the convergence rate becomes almost linear.

This convergence pattern was similar for all the tested cases. The calculation of the Jacobian at every iteration is important in speeding up the convergence but, on the other hand, calculating the Jacobian at every iteration is time consuming and

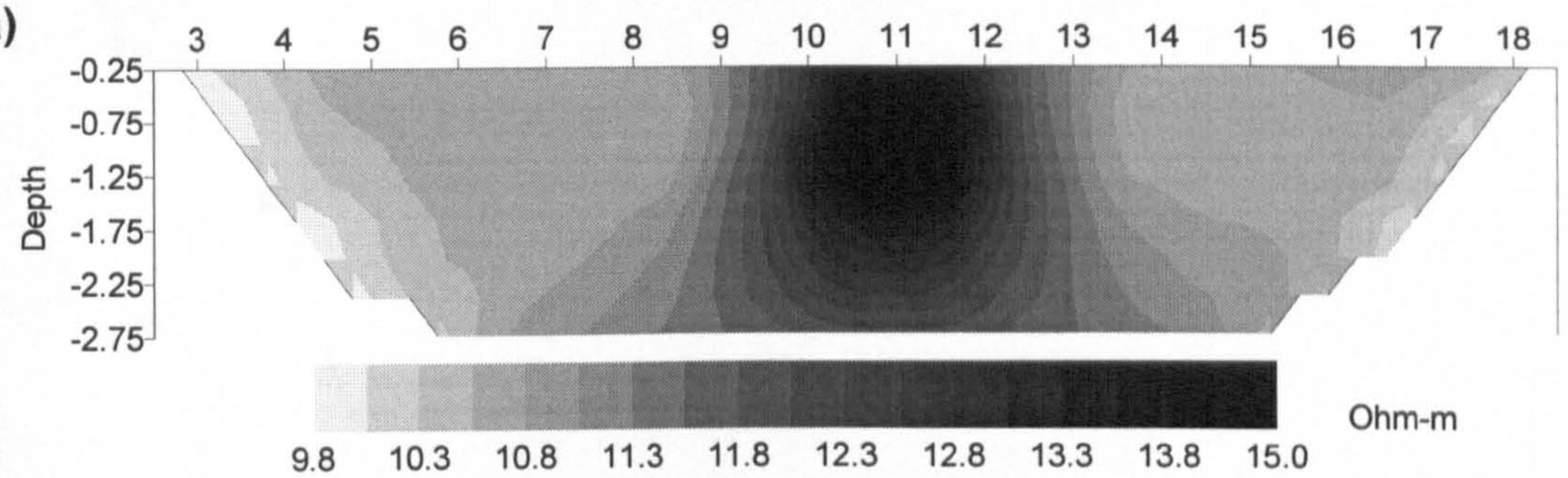


Figure 5.10: Reconstruction of Model 1 (dipole-dipole data): a) generalized BP (1 iteration), b) generalized BP (10 iterations).

therefore the Jacobian is updated only during the first iterations where the RMS drops fast (indicating in this way sizable changes in the resistivity distribution) and afterwards only every other 3 iterations. This reduces the time that the reconstruction takes without significantly affecting the convergence. A simplified flow-chart of the algorithm is shown in figure 5.12.

In Figure 5.13 the reconstruction results using several known arrays are shown. The reconstruction of the Wenner and pole-dipole data sets for Model 1 are depicted in Figures 5.13a,b. In Figure 5.13c the reconstruction of Model 2 for pole-pole data is shown The results indicate that the technique can be used for reconstructing data sets obtained from any array. For all reconstructions only the positive entries of the Jacobian matrix were considered and the additive BP formula was used (equation 5.7). Note that for the pole-dipole array the full data set was considered (both A, M-N and N-M, A arrangements) since using only the data set for the arrangement (A,M-N) results in an asymmetric reconstruction.

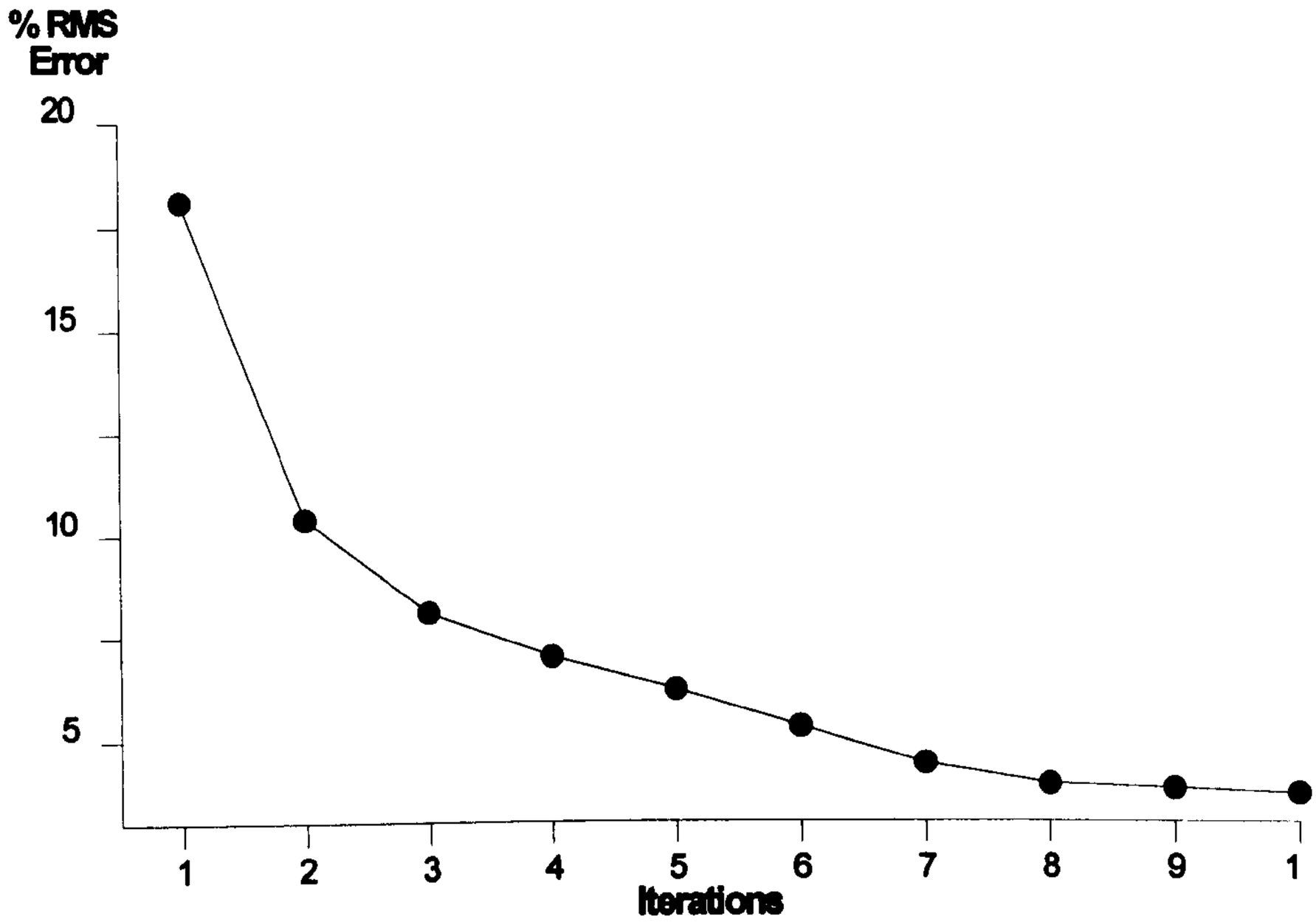


Figure 5.11: The RMS error curve for the reconstruction of the dipole-dipole set for Model 1.

Several other tests were conducted with a variety of synthetic models. In Figure 5.14 the reconstructions for the dipole-dipole modelled data from Models 2, 3, 4 are shown. In all of the examples the maximum dipole separation was six dipole lengths (n=6) and the entire measuring pattern included 20 measuring probes. The data sets were corrupted with 5% Gaussian noise.

The results indicate that the algorithm can give reliable reconstructions which delineate the subsurface structures fairly well in both a quantitative and a qualitative fashion. Further, the algorithm behaved well with noisy data and it still retained its stability. No major artifacts appeared which could have indicated an inherent problem of the algorithm towards noisy data. Stability problems appeared only when the level of the noise was very high.

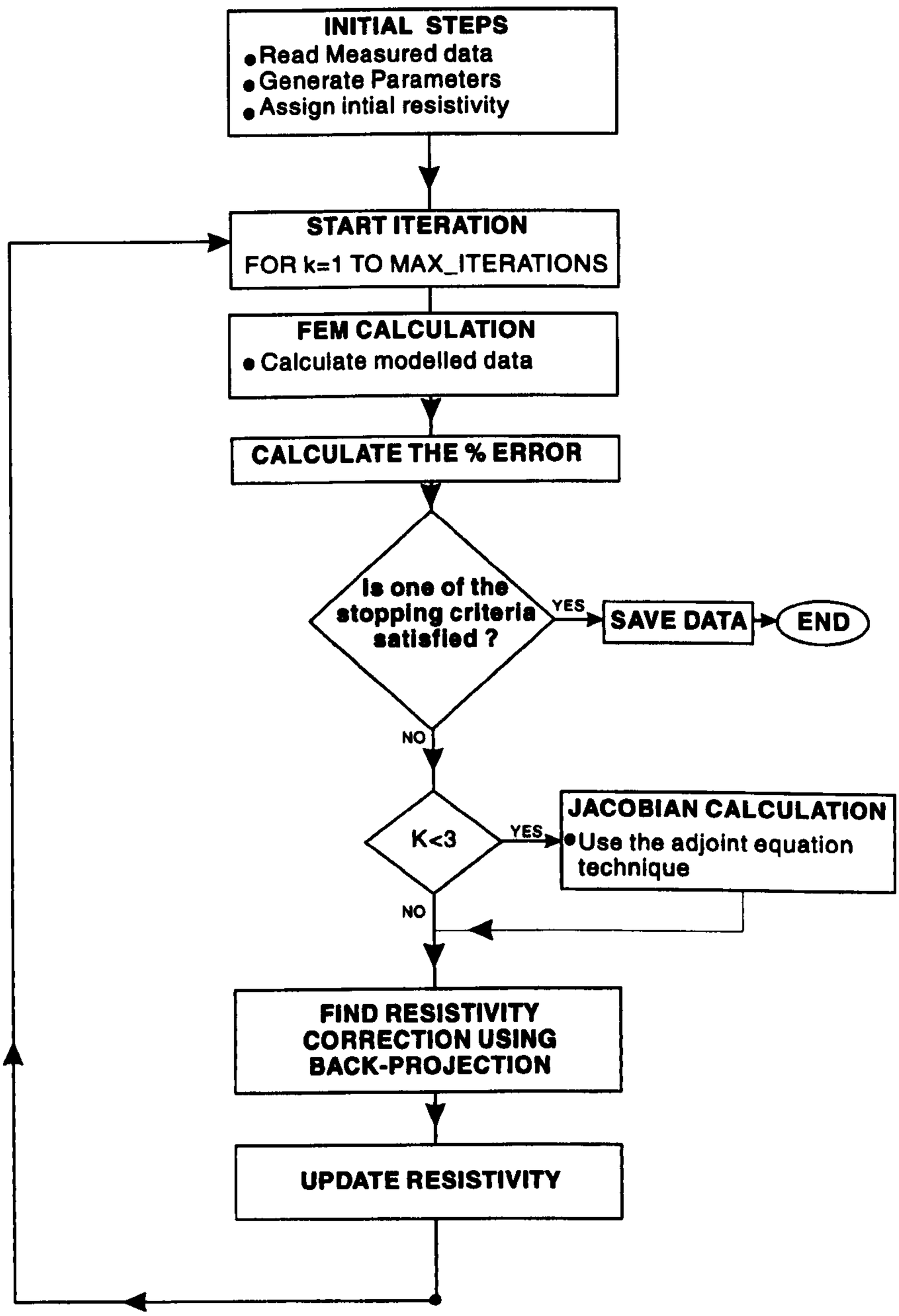


Figure 5.12: A simplified flow-chart of the algorithm.

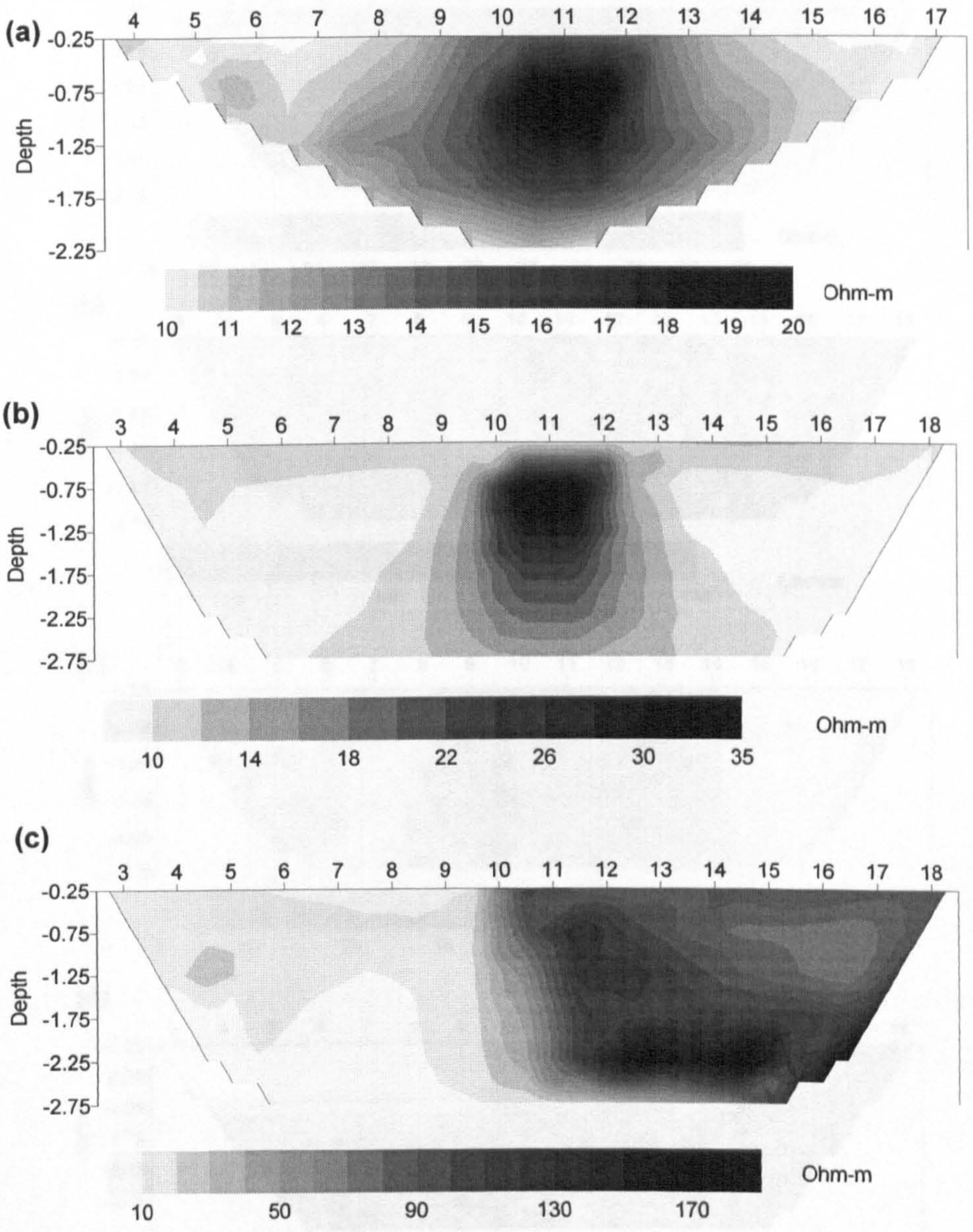


Figure 5.13: Reconstruction of Model 1 using the generalized BP scheme: a) Wenner data set (10 iterations, 2.8% RMS error), b) pole-dipole data set (11 iterations, 3.5% RMS error). c) Reconstruction of Model 2 using the generalized BP scheme for pole-pole data (8 iterations, 4.2% RMS error).

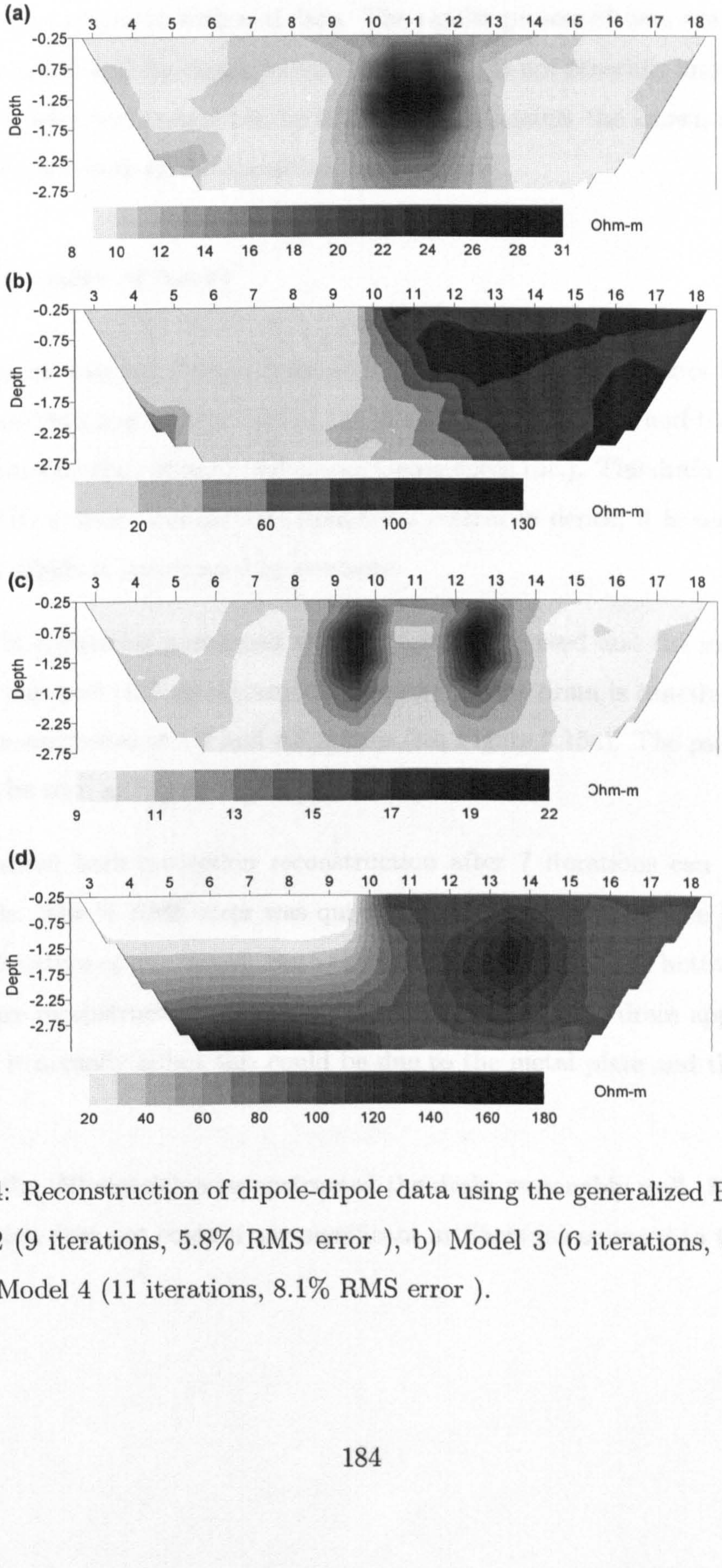


Figure 5.14: Reconstruction of dipole-dipole data using the generalized BP scheme: a) Model 2 (9 iterations, 5.8% RMS error), b) Model 3 (6 iterations, 6.3% RMS error), c) Model 4 (11 iterations, 8.1% RMS error).

5.2.4 Real data reconstructions

The algorithm was tested with real data. The results presented here are from areas where the position and the characteristics of the targets are generally known. Therefore, the reconstructed images can be tested directly against the known targets and an objective validation of the algorithm can be made.

Drain (University of York)

A dipole-dipole data set from a drain situated outside the Electronics building at University of York was collected using the Sting resistivity meter and the Swift automatic multi-electrode system (Advanced Geosciences Inc.). The drain has surface dimensions 60 x 60cm and extends from 0-1.2 metres in depth; it is sealed with a metal cover which is surrounded by concrete.

A total of 24 electrodes positioned at 60cm apart were used and the maximum n-separation was n=8 (137 measurements). Note that the drain is exactly positioned between the electrodes at 4.2 and 4.8 metres (see Figure 5.15a). The pseudosection results can be seen in Figure 5.15b.

The generalized back-projection reconstruction after 7 iterations can be seen in Figure 5.15c. The % RMS error was quite high 14.3% but this can be justified by the full 3-D nature of the target. Both the lateral location and the bottom limits of the drain are reconstructed accurately. The upper limit of the drain appears to be lower than it actually is but this could be due to the metal plate and the concrete on the top.

In general the BP algorithm reconstructed the drain reasonably well. Further, the reconstruction does not contain any significant artifacts (as opposed to the pseudo-section).

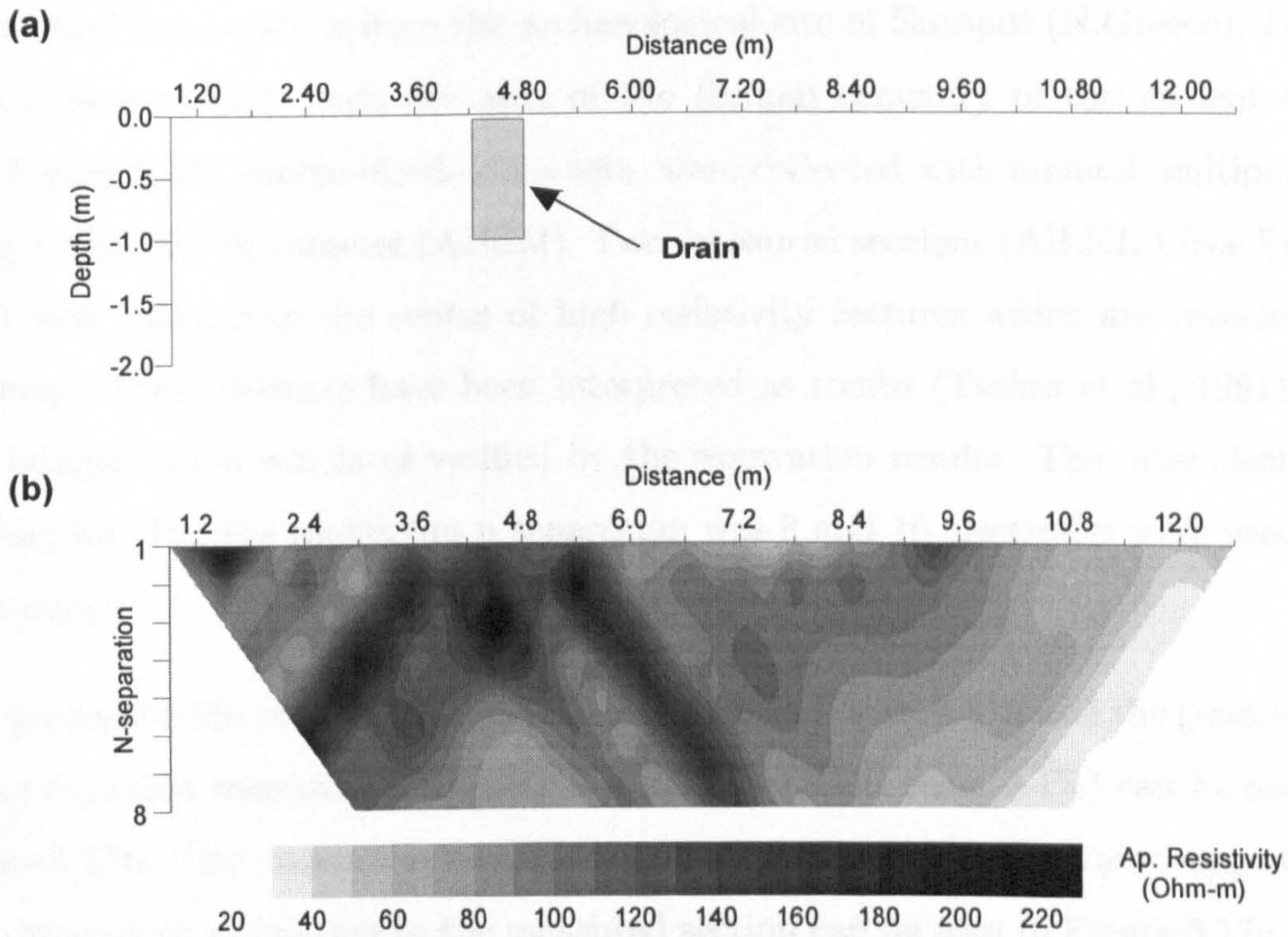


Figure 5.15: Reconstruction of dipole-dipole data measured over a drain (University of York): a) the exact location of the drain in relation to the measured section, b) the measured data set in a pseudosection form, c) reconstruction using the generalized BP algorithm (7 iterations, 14% RMS error).

Roman cemetery (Europos, N. Greece)

The second case study is from the archaeological site of Europos (N.Greece). Based on a resistivity map from the area of the Roman cemetery of the ancient town (see Figure 5.16), dipole-dipole data sets, were collected with manual multiplexing using a SAS 300 Terrameter (ABEM). Two measured sections (AB,KL) (see Figure 5.16) were taken over the centre of high resistivity features which are revealed on the map. These features have been interpreted as tombs (Tsokas et al., 1991) and this interpretation was later verified by the excavation results. The inter-electrode spacing was 1m, the maximum n separation was 8 and 16 electrodes were used (76 data-points).

The pseudosection of the data set AB is depicted in Figure 5.17a and the generalized back-projection reconstruction after 9 iterations (RMS error 9.4%) can be seen in Figure 5.17b. The excavation revealed a rectangular grave: its approximate position and dimensions in relation to the measured section can be seen in Figure 5.17c. The BP reconstruction is in good agreement with the excavation results.

In Figure 5.18a the pseudosection of the data set KL is shown. Note that the pseudo-section image is quite complicated. The generalized back-projection reconstruction after 11 iterations (RMS error 12.2%) can be seen in Figure 5.18b. The excavation revealed a grave with a collapsed roof: its approximate position and dimensions in relation to the measured section can be seen in Figure 5.18c. The reconstruction produced useful information about the shape and the location of the anomaly (the side walls of the tomb can be distinguished), and they are in agreement with the excavation results.

Fountains Abbey, (N. Yorkshire)

Two case studies from the Fountains Abbey World Heritage site are presented here.

The fist case study is from the guest hall area. A detailed twin-probe profiling

Figure 5.16: Resistivity contour map obtained from the area of the ancient cemetery of Europos. The observed resistivity anomalies have been interpreted as tombs (Tsokas et al., 1991). The measured sections (AB), (KL) are shown as well.

survey at the site revealed the foundations of the guest hall at the Western side of the Abbey. The survey results have been presented in Figure 2.13 [section 2.3.2] and the plan of the guest-hall after the interpretation of the resistivity survey is depicted in Figure 5.19a.

A dipole-dipole data set was obtained over the section F1 (the section is also depicted in Figure 5.19a) with a South-to-North direction. A total of 65 electrodes positioned 0.5m apart were used and the maximum n-separation was n=7 (413 measurements)[4]. The pseudosection results are shown in Figure 5.19b and the BP reconstruction (11 iterations, 7.5% RMS error) is shown in Figure 5.19c. The reconstructed image produced two major resistive features which coincide exactly with the location of the wall foundations. Further, a less resistive region situated at the middle of the section can be due to the side-effect of the pillar foundations which are located quite close to the measuring section (20cm). The reconstruction positioned the upper limits of the walls at a depth of approximately 50cm. This depth is probably unrealistic - the real depth should be around 20-30cm (Emerick, 1995, personal communication).

The second case study is from the area of the mill of the Abbey. A drain runs across a narrow strip of land which is enclosed by the river Skell. A dipole-dipole data set (24 electrodes 1m apart, maximum n=7) was obtained over a section situated orthogonally to the drain. The edge of the drain is visible and its *exact* position in relation to the measured section is depicted in Figure 5.20a.

The pseudosection of the data set is depicted in Figure 5.20b - note that the pseudosection slightly misplaces the target to the left. The generalized back-projection reconstruction after 9 iterations (RMS error 6.3%) can be seen in Figure 5.20c. The BP results are in very good agreement with the reality - both the position and the size of the drain is successfully reconstructed.

---

[4]The Sting resistivity meter and the Swift automatic multi-electrode system (Advanced Geosciences Inc.) were used.

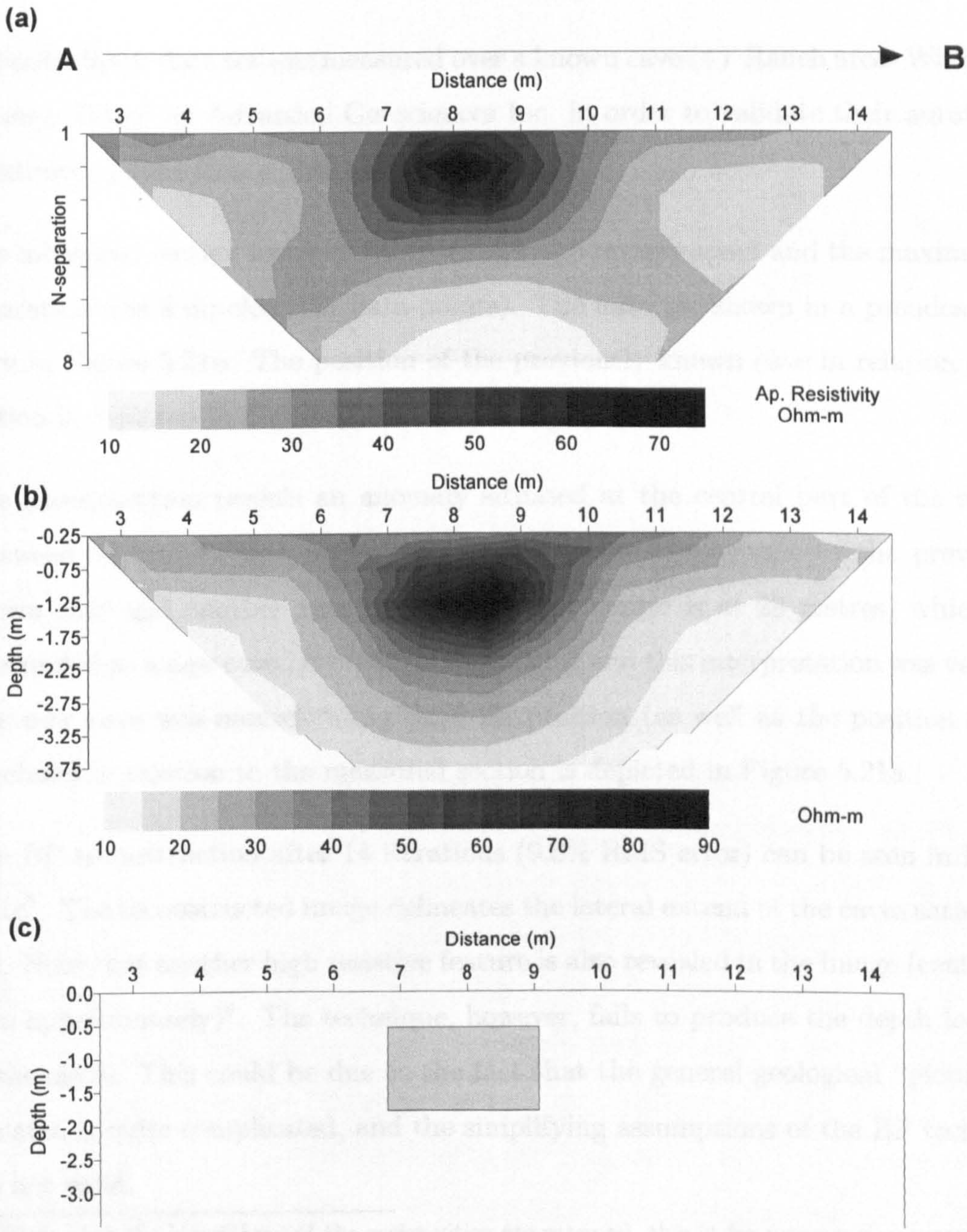


Figure 5.17: Reconstruction of the dipole-dipole data measured over a tomb at the roman cemetery of Europos (N. Greece) (section AB): a) the measured data set in a pseudosection form, b) reconstruction using the generalized BP algorithm (9 iterations, 9.4% RMS error), a) the approximate location of the excavated tomb in relation to the measured section.

"Sting" Cave (Williamson County, Texas)

A dipole-dipole data set was measured over a known cave (4T Ranch area- Williamson County, Texas) by Advanced Geosciences Inc. in order to validate their automated resistivity system (Sting/Swift system).

The measured section involves 28 electrodes, 4.5 metres apart and the maximum n-separation was 8 dipoles (171 data-points). The data are shown in a pseudosection form in Figure 5.21b. The position of the previously known cave in relation to the section is depicted in Figure 5.21a.

The pseudosection reveals an anomaly situated at the central part of the section (between 62 and 71 metres approximately) which corresponds to the previously known cave and another resistive anomaly (its centre is at 28 metres) which was interpreted as a new cave. A borehole was drilled and this interpretation was verified. The new cave was named "sting" and its position (as well as the position of the borehole) in relation to the measured section is depicted in Figure 5.21a.

The BP reconstruction after 14 iterations (9.8% RMS error) can be seen in Figure 5.21c[5]. The reconstructed image delineates the lateral extend of the caves satisfactorily. Note that another high resistive feature is also revealed in the image (centred at 98m approximately)[6]. The technique, however, fails to produce the depth location of the caves. This could be due to the fact that the general geological "picture" of the area is quite complicated, and the simplifying assumptions of the BP technique are not valid.

---

[5]Note that the logarithms of the resistivities are mapped- this is for presentation reasons only when the resistivity contrasts are very high.

[6]Further processing of the data -it will be presented in the next chapter- makes us believe that this feature is probably another cave

5.2.5 Discussion

The described algorithm, although by no means theoretically correct, is physically reasonable. The back-projection procedure can be viewed as a pseudo-inverse operation: following Shima (1992) if $\mathbf{J}$ is the Jacobian matrix, the back-projection approximates $\mathbf{J^{-1}}$ by $\mathbf{WJ^T}$ where $\mathbf{W}$ is a diagonal matrix whose elements are the inverse of the sum of the sensitivities related to each measurement.

Because of its approximate nature the convergence of the algorithm is not guaranteed, nor can the speed of the convergence be predicted. Divergence may sometimes occur when the data noise is high. The convergence, if there is convergence at all, can be speeded up by the use of an over-relaxation factor within the reconstruction algorithm. In most of the tested cases the algorithm presented a stable convergence behaviour and usually took less than 10 iterations to reach to a minimum error.

Barker (1992) referred to possible inherent problems with simple back-projection methods when near-surface lateral changes in resistivity occur (especially when arrays sensitive to such types of resistivity anomalies are used) and he suggested that these effects can be propagated and projected deeper, hence creating artifacts. This can be true in the case of simple graphical back-projection schemes such as Bristow's method, although it should be borne in mind that full back-projection is a summation procedure so that the correction factor for each parameter is influenced by all of the measurements that affect this parameter. Therefore, the effect of an individual measurement need not to be decisive for determining any particular resistivity. Moreover, in the current scheme, the existence of the weighting terms are another factor which limit the propagation of the lateral effects with depth is limited. The results presented here verify this observation.

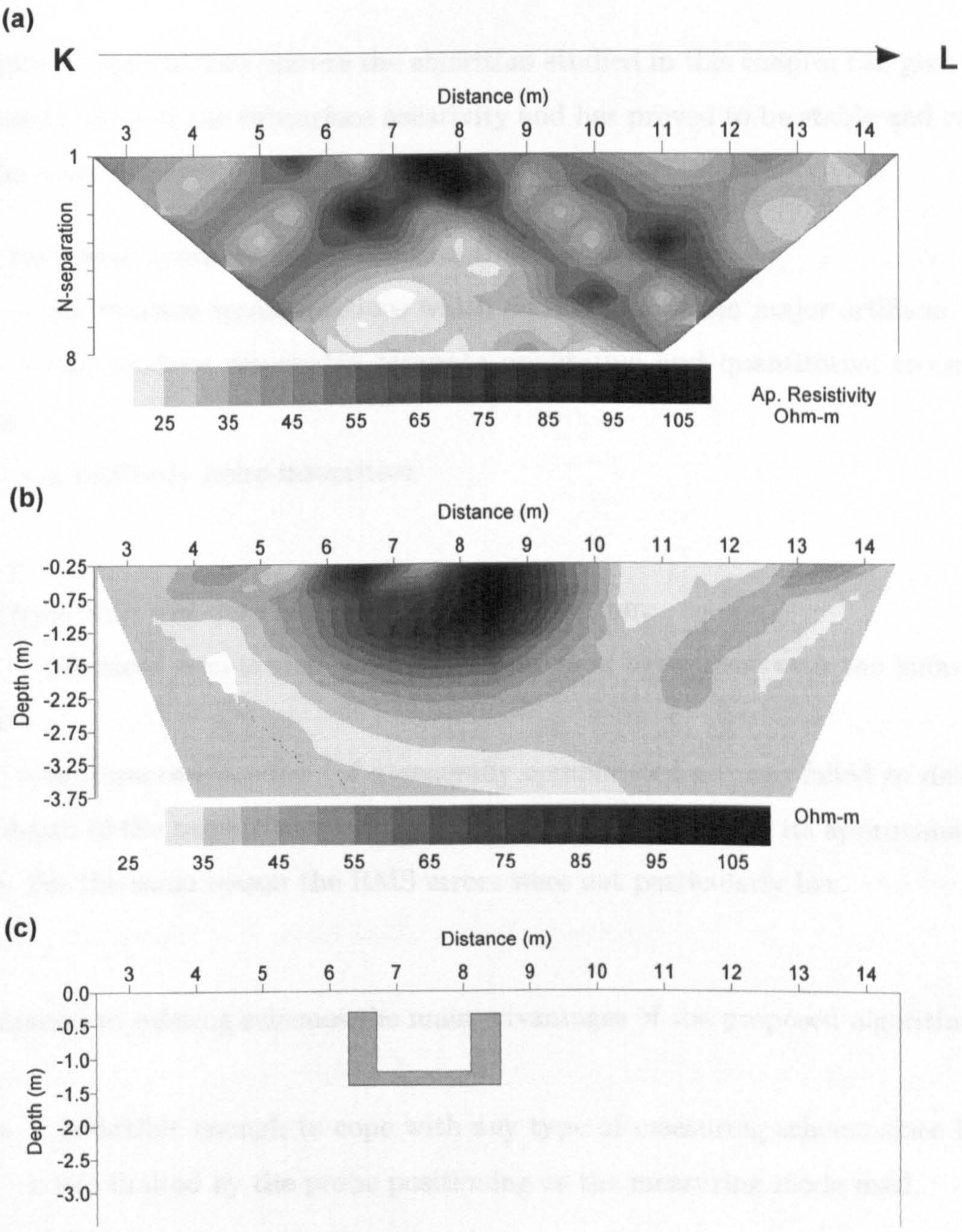


Figure 5.18: Reconstruction of the dipole-dipole data measured over a tomb at the roman cemetery of Europos (N. Greece) (section KL): a) the measured data set in a pseudosection form, b) reconstruction using the generalized BP algorithm (11 iterations, 12.2% RMS error), a) the approximate location of the excavated tomb in relation to the measured section.

5.2.6 Conclusions

Despite its approximate nature the algorithm studied in this chapter can give useful reconstructions of the subsurface resistivity and has proved to be stable and reliable in the cases tested.

The tests with synthetic data indicated that the algorithm:

- can produce reconstructions which do not suffer from major artifacts.
- can produce reasonably accurate qualitative and quantitative reconstructions.
- is relatively noise-insensitive.

The tests with real data indicated that the algorithm:

- produced results that are generally in good agreement with the known targets.
- in some case-studies (of a generally complicated nature) failed to delineate the depth of the targets successfully. This problems are due to its approximate nature. For the same reason the RMS errors were not particularly low.

Compared to existing schemes the main advantages of the proposed algorithm are:

- It is flexible enough to cope with any type of measuring scheme since its use is not limited by the probe positioning or the measuring mode used.
- It can give reconstructions which produce both qualitative and quantitative information of the resistivity distribution - a major advantage over the non-iterative back-projection schemes which can only give qualitative information.
- It is computationally fast compared to typical non-linear schemes since it avoids the matrix inversion procedure. Note that the matrix inversion procedure when very large data sets (i.e. fully tomographic) are considered is

extremely time consuming.

- It can be modified to include most of the known approximate reconstruction algorithms.

Overall, it is believed that the algorithm can serve as a practical tool for the preliminary interpretation of tomographic resistivity data. Moreover, the reconstructed images can be used as initial estimates for accurate inversion schemes.

5.3 Chapter Overview

In this Chapter approximate inversion techniques applied to the 2-D reconstruction of earth resistivity data were presented.

In the first part of this chapter widely used approximate reconstruction techniques such as the pseudosection, Barker's method, and back-projection algorithms have been presented. It was shown that the Jacobian matrix can be used to explain why these techniques are effective. Reconstruction with synthetic data tests indicated some of the merits and demerits of the tested techniques.

On the basis of the theoretical characteristic of the techniques a generalized back-projection algorithm was proposed. The algorithm can include many of the existing approximate algorithms and can cope with any type of measuring scheme. This is achieved by recognizing the major importance of the Jacobian matrix within the approximate reconstruction procedure.

The features of the algorithm were presented in detail and extended tests with synthetic data were conducted. Finally, tests of the algorithm with real data were presented. Despite its approximate nature the algorithm proved to be reliable and noise insensitive and produced good quality reconstructions. The tests with real data indicated that it can be a useful tool for data interpretation.

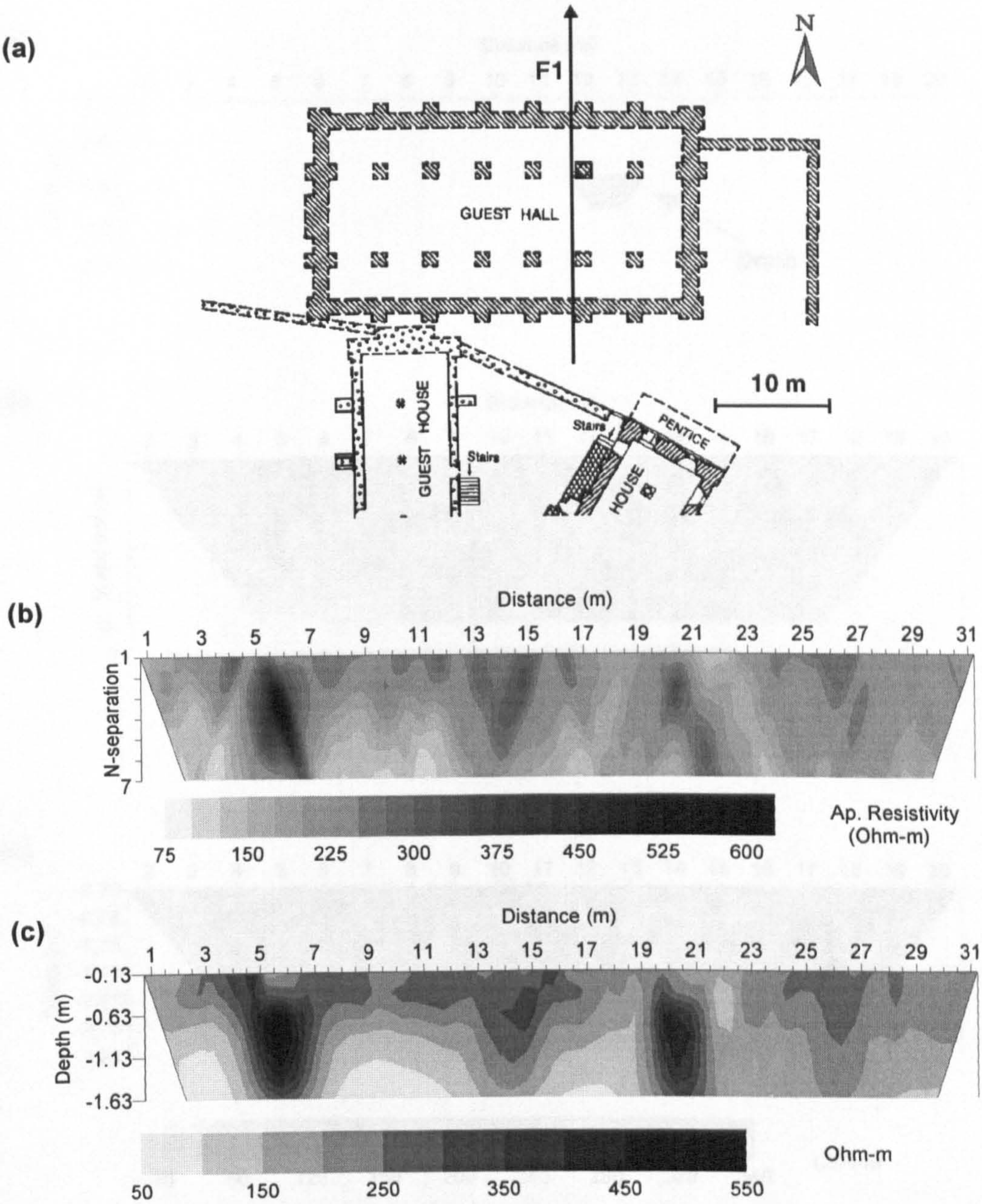


Figure 5.19: Reconstruction of the dipole-dipole data measured over the area of the Guest-Hall at Fountains Abbey (N. Yorkshire): a) the location of the section in relation to the Hall's foundations, b) the measured data set in a pseudosection form, c) reconstruction using the generalized BP algorithm (11 iterations, 7.5% RMS error).

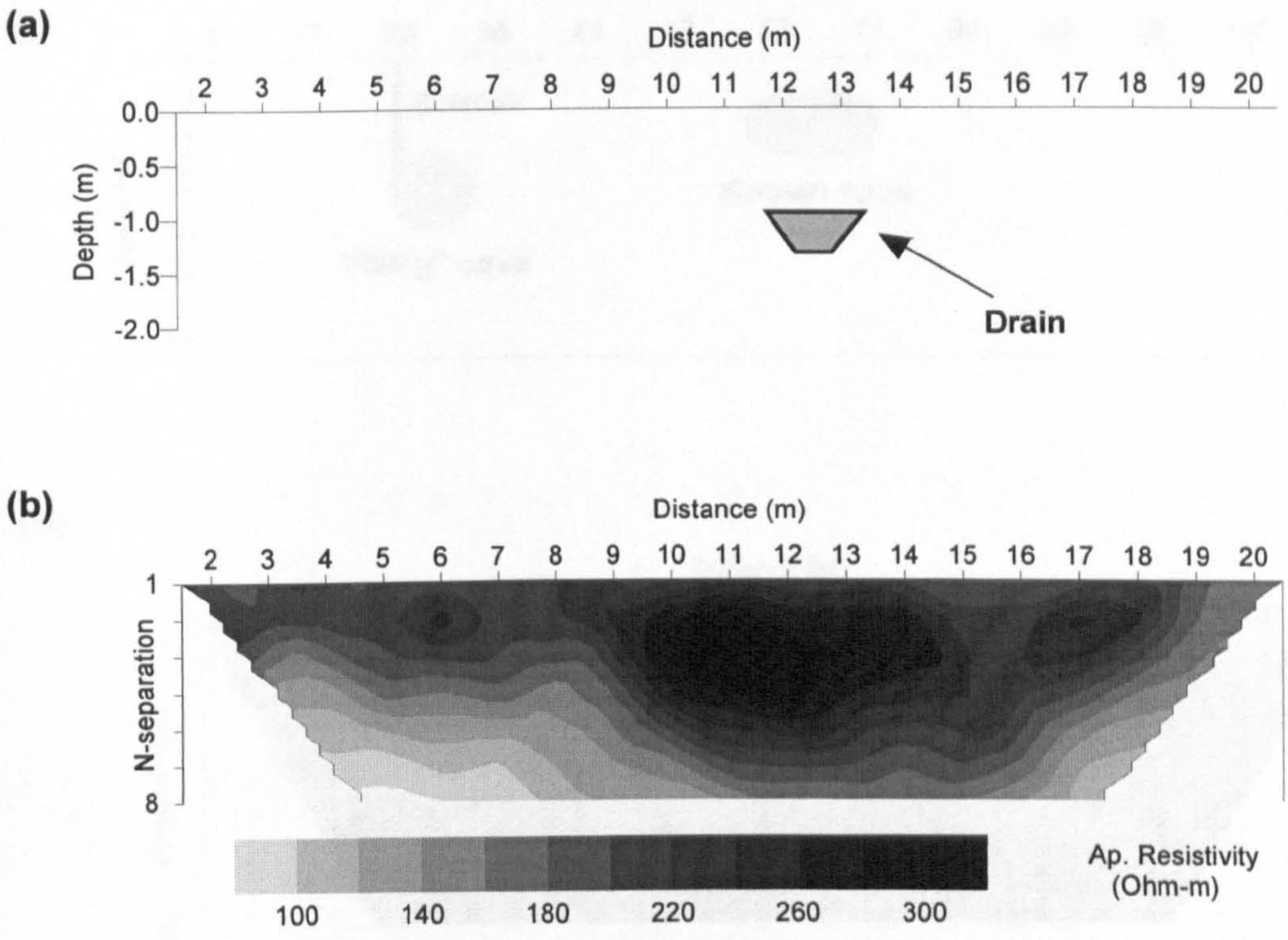


Figure 5.20: Reconstruction of the dipole-dipole data measured over a drain at the Fountains Abbey (N. Yorkshire): a) the exact location of the drain in relation to the measured section, b) the measured data set in a pseudosection form, c) reconstruction using the generalized BP algorithm (9 iterations, 6.3% RMS error).

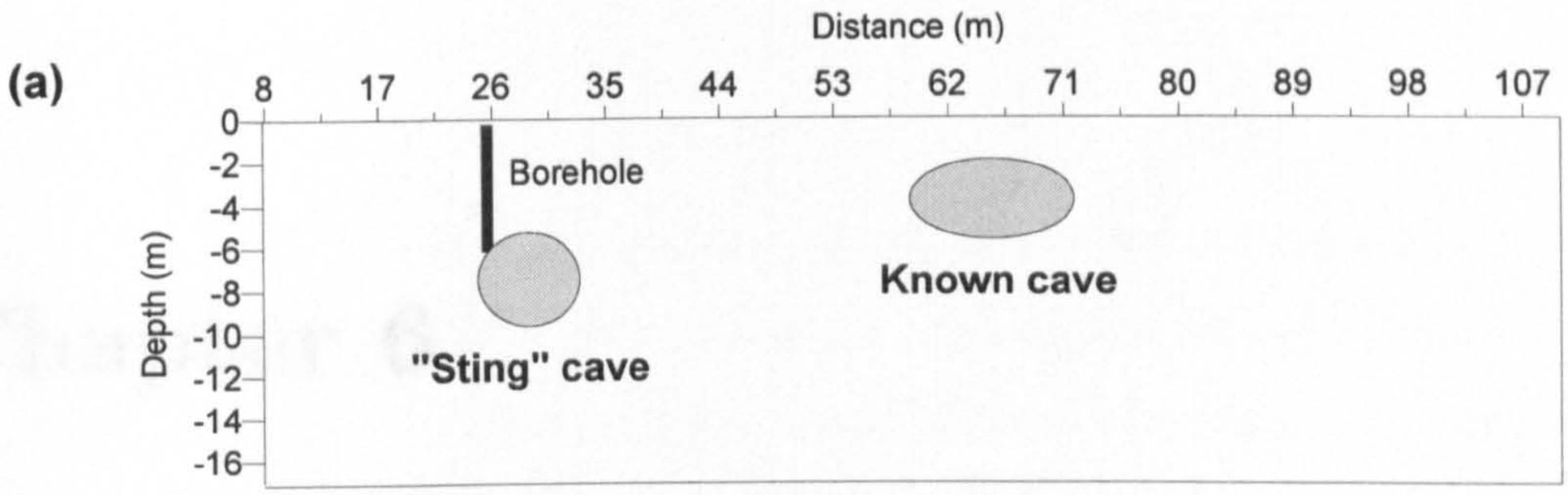


Figure 5.21: Reconstruction of the dipole-dipole data measured over a cave ("Sting" Cave, Williamson County, Texas): a) the exact location of the known caves in relation to the measured section, b) the measured data set in a pseudosection form, c) reconstruction using the generalized BP algorithm (14 iterations, 9.8% RMS error).

Chapter 6

2-D Inversion of Earth Resistivity Data Using Accurate Inversion Techniques

In this Chapter non-linear inversion techniques applied to the 2-D reconstruction of earth resistivity data will be presented. In the first part of this chapter widely used techniques such as the non-linear least-squares method, Marquadt's method, and smoothness constrained (Occam) inversion will be reviewed. Further, approaches such as inversion with prior information, simulated annealing and maximum entropy will be presented as well. The advantages and limitations of the techniques will be demonstrated by means of synthetic examples.

In the second part of this chapter a fast smoothness constrained inversion algorithm which uses a quasi-Newton technique for updating the Jacobian matrix is proposed. The need for such an algorithm is discussed. The features of the algorithm are presented in detail and comparisons to other techniques will be shown. Finally, tests of the algorithm with real data will be presented.

6.1 A review of existing non-linear inversion techniques

In this section non-linear techniques for solving the resistivity inverse problem are presented. These techniques are: non-linear least-squares, weighted least-squares, Marquadt's method, smoothness constrained (Occam) inversion, inversion with *a priori* information, maximum entropy, and simulated annealing. The relative merits and demerits of the techniques are discussed.

Notation

In this section (and throughout this chapter) the following general notation is used (The quantities corresponding to the resistivity case are specified in parenthesis) .

y is the m-dimensional observation vector (apparent resistivities).

x is the n-dimensional model parameter vector (resistivities).

J is the m x n Jacobian matrix.

f(x) (apparent resistivities) is the m-dimensional measurement vector which corresponds to the model **x**.

f represents the forward modelling procedure (FEM).

dy is the m-dimensional vector of the differences (error) between the observed and modelled data.

Tested models

Most of the techniques presented in this section were tested using a series of synthetic data sets. For all of the tested techniques, an identical four layer parametrization scheme was followed. In the main, each parameter is a square block with side equal to the inter-electrode spacing, but the parameters at the side and bottom borders are set to be five times larger in one dimension (see Figure 6.1a).

Three of the models which were used to produce the synthetic data are depicted in Figures 6.1b,c,d. We will be referring to those models as Model 1 (two resistive prisms), Model 2 (vertical discontinuity with overburden), Model 3 (complicated

structure).

6.1.1 Least-squares inversion

The least-squares method is one of the most popular techniques for solving optimization problems. Extended reviews of the technique can be found in many works (e.g. Lowson and Hanson, 1974; Press et al., 1987) and so only a brief description of the method is presented here. The least-squares technique has been used extensively for the solution of a wide range of geophysical optimization problems (Jupp and Vozoff, 1975; Lines and Treitel, 1984).

The resistivity non-linear problem can be described as (see equation 2.7):

$$\mathbf{f(x) = y}$$

where $\mathbf{y}$ is the vector of observed measurements, $\mathbf{x}$ is the vector describing the unknown parameters, and $\mathbf{f(x)}$ is the forward model solution which is a known function of the parameter distribution. As described in section 4.1, expanding in a Taylor series and ignoring the higher-order terms yields:

$$\mathbf{f(x + dx)} \simeq \mathbf{f(x) + Jdx} \tag{6.1}$$

where $\mathbf{dx}$ is the resistivity correction vector and $\mathbf{J}$ is the Jacobian matrix. The general optimization procedure seeks to minimize the errors between the observed and predicted data. The least-squares approach seeks to find the optimum resistivity correction vector $\mathbf{dx}$ for which the square sum of errors $\mathrm{S(x)}$ becomes minimum, where $\mathrm{S(x)}$ is

$$S(\mathbf{x}) = \mathbf{dy}^T\,\mathbf{dy} = (\mathbf{y} - \mathbf{f(x + dx)})^T\,(\mathbf{y} - \mathbf{f(x + dx)}) \tag{6.2}$$

This is achieved by setting the derivative of $\mathrm{S(x)}$ in respect of $\mathbf{dx}$ to zero [1].

$$\frac{\partial S(\mathbf{x})}{\partial \mathbf{dx}} = 0 \tag{6.3}$$

[1] As Lines and Treitel (1984) emphasized, another criterion that should hold is that the second derivative of $\mathrm{S(x)}$ should be positive definite. They showed that this holds when the errors are linear functions of the parameters, and that otherwise minimization does not necessarily occur.

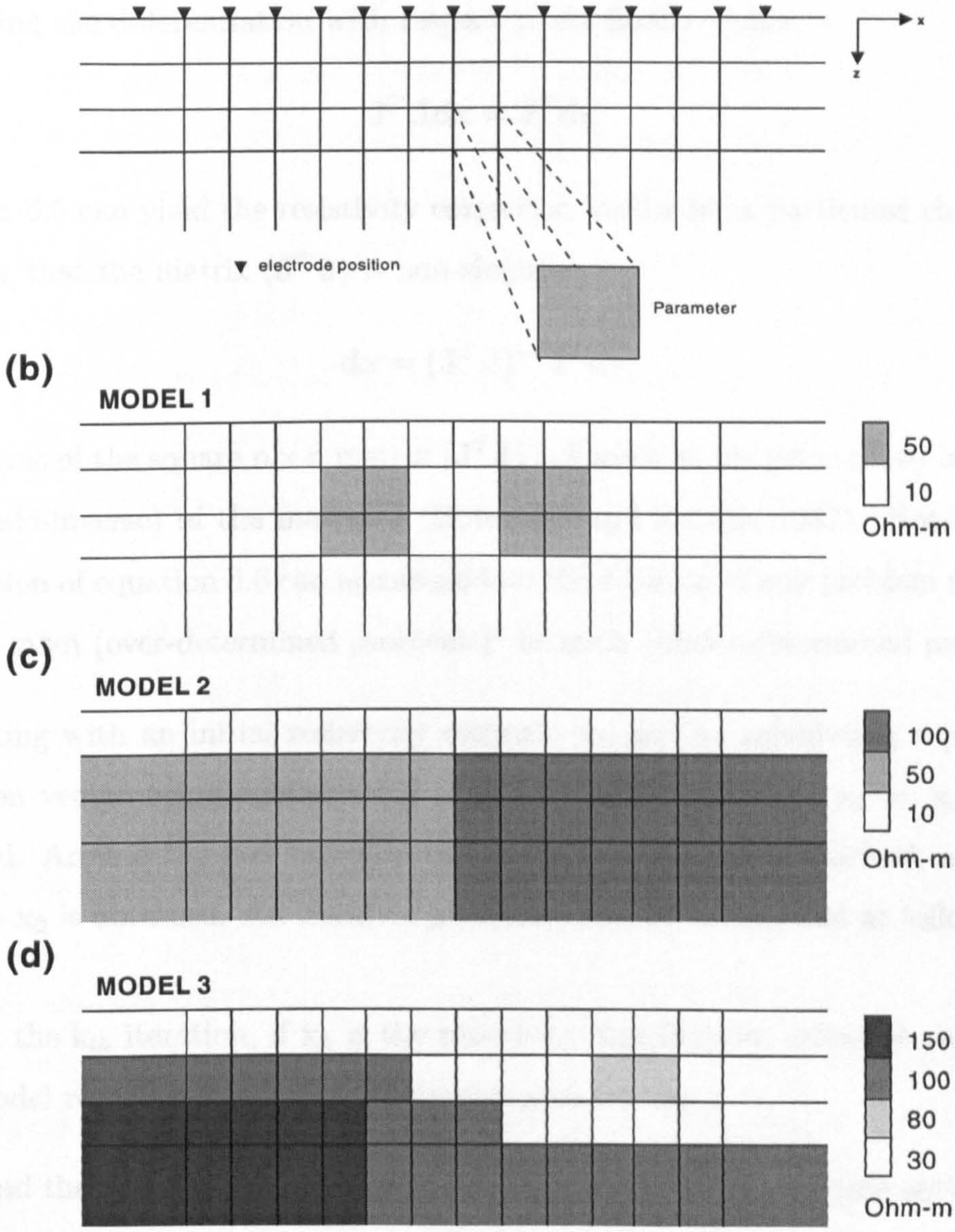


Figure 6.1: a) The parametrization scheme used in this section. The synthetic models used in this section: b) Model 1 (two resistive prisms), c) Model 2 (vertical discontinuity with overburden), d) Model 3 (complicated structure).

Due to equations 6.1 and 6.2 equation 6.3 becomes:

$$\frac{\partial S(\mathbf{x})}{\partial \mathbf{dx}} = \frac{\partial[(\mathbf{y} - \mathbf{f(x)} - \mathbf{Jdx})^T(\mathbf{y} - \mathbf{f(x)} - \mathbf{Jdx})]}{\partial \mathbf{dx}} = 0 \tag{6.4}$$

Performing the differentiation with respect to **dx** finally yields:

$$\mathbf{J}^T\mathbf{Jdx} = \mathbf{J}^T\mathbf{dy} \tag{6.5}$$

Equation 6.5 can yield the resistivity correction vector for a particular choice of **x**, assuming that the matrix $(\mathbf{J}^T\mathbf{J})$ is non-singular:

$$\mathbf{dx} = (\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T\mathbf{dy} \tag{6.6}$$

The inverse of the square n x n matrix $(\mathbf{J}^T\mathbf{J})$ is known as the generalized inverse (or the pseudo-inverse) of the matrix **J** (Hohmann and Raiche, 1987). Note that the formulation of equation 6.6 can accommodate the solution of any problem regardless whether m>n (over-determined problems)[2] or m<n (under-determined problems).

By starting with an initial resistivity estimate $\mathbf{x}_o$ and by calculating a parameter correction vector using equation 6.6 a new resistivity estimate $\mathbf{x}_1 = \mathbf{x}_o + \mathbf{dx}$ is produced. Around the new estimate the model can again be linearized, and a new estimate $\mathbf{x}_2$ is obtained. An iterative procedure can be established as follows:

- At the k_{th} iteration, if $\mathbf{x}_k$ is the resistivity distribution, calculate the forward model response $\mathbf{f}(\mathbf{x}_k)$ and the Jacobian matrix $\mathbf{J}_k$
- Find the parameter correction vector $\mathbf{dx}_k$ which is given by (see equation 6.6)

$$\mathbf{dx}_k = (\mathbf{J}_k^T\mathbf{J}_k)^{-1}\mathbf{J}_k^T\mathbf{dy}_k \tag{6.7}$$

  where $\mathbf{dy}_k = \mathbf{y} - \mathbf{f}(\mathbf{x}_k)$.
- The new resistivity estimate $\mathbf{x}_{k+1}$ is given by

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{dx}_k \tag{6.8}$$

[2]Note that most of the resistivity inverse problems are over-determined by design.

- The iterative procedure is repeated until a stopping criterion is met (i.e. no significant decrease in error is achieved with further iterations.)

This procedure is known as non-linear least-squares minimization or the Gauss-Newton method. Note that the initial problem of solving the system of non-linear equations has been changed to that of solving a system of linear equations at every iteration step. Non-linear least-squares methods have been reported to be quite sensitive into the choice of the initial parameter vector $\mathbf{x}_o$ and as suggested by Box and Kanemasu (1972), if the initial parameter choice is poor it has sometimes been found that severe oscillations of the solution vector occur through the iterations. If, for example, the initial property choice is too far away from the true solution, the correction vector **dx** may become too large and hence the assumption of linearity becomes invalid [3], leading to divergence. Smith and Vozzoff (1984) also suggested that sometimes the linearity assumption causes the procedure to choose the local minimum nearest to the initial model, hence rendering the results dependent on the initial parameter choice.

Ill-conditioning and data error

As far as the resistivity inversion is concerned, a quite significant problem arises from the fact that the matrix $(\mathbf{J}^T\,\mathbf{J})$ is near-singular, a direct consequence of the reportedly ill-conditioned nature of the resistivity problem. Lanczos (1960) discussed the physical meaning of nearly-singular systems in relation to the observation error. He showed that the existence of small eigenvalues signifies that there are certain linear combinations of the unknown parameters which are only weakly represented in the system. In physical terms this means that there are some relatively unimportant parameters (see section 4.4). Therefore, it is a lack of useful information which creates the ill-conditioning and, although there are ways to handle ill-conditioning (discussed below) the base reasons need to be borne in mind. To quote Lanczos

[3]The Taylor expansion procedure assumes that **dx** is "small" enough.

(1960):

> ...there is an impression that an ill-conditioned matrix is merely in a mathematical condition which could be remedied by the proper know-how. In actual fact we should recognize that a lack of information cannot be remedied by any mathematical trickery...

If the observations were completely error free the inversion of the near-singular matrix would not be a problem, however this is not the case, and it is the combination of the ill-conditioning with the observation errors that creates the problem. Lanczos (1960) showed that for an idealized linear system (m=n) the amplitude of the error in the solution x_i^e associated with a particular measurement error y_i^e and the respective eigenvalue λ_i is given by the following relation:

$$|x_i^e| = \frac{1}{\lambda_i}|y_i^e| \tag{6.9}$$

An eigenvalue of 0.01 and a measurement error of 1% will magnify the error in the solution vector by a factor of 100. The solution of a near-singular system will result into unreliable solutions not only for the parameters which are weakly represented in the system but also for the parameters which are strongly represented in the system[4]

Lanczos's observations for square linear systems can be extended into the non-linear procedure. Treating the system of equation 6.7 using direct matrix inversion will result into unacceptable solutions for the resistivity inverse problems. Alternative techniques have to be used in order to tackle this problem and some of them will be presented in the following subsections.

---

[4]Note that for the resistivity case the observation errors are typically larger than 3% and a typical conditioning number is $10^3 - 10^4$

6.1.2 Weighted least-squares

In this part a variation of the least-squares technique called weighted least-squares is presented. Weighted least-squares does not improve the conditioning of the inverse problem considerably, but can take into account the observation errors as well as other physical constraints (i.e. positive properties). They are presented at this point simply because similar weighting principles can be used within the framework of all the techniques presented in this section.

Error weighting

It has already been shown that the ill-conditioning of the resistivity problem when combined with the limited accuracy of the measurements, renders the solution of the inverse problem unstable. In order to take the measurement accuracy into account a weighted least-squares solution can be formed.

In particular, instead of minimizing the sum of square errors an alternative cost function $S(\mathbf{x})_w$ can be minimized (Lawson and Hanson, 1974):

$$S(\mathbf{x})_w = (\mathbf{W}\,\mathbf{dy})^T (\mathbf{W}\,\mathbf{dy}) = \mathbf{W^2}\,\mathbf{dy}^T\,\mathbf{dy} = \mathbf{W^2} S(\mathbf{x}) \tag{6.10}$$

where S(**x**) is the least-squares cost function, (see equation 6.2) and **W** is a m x m diagonal matrix whose elements are weights which depend on the accuracy of each particular data point. One systematic way to assign weights is to use the standard deviation, σ information for each measurement y_i. A diagonal element of the **W** matrix is then defined as:

$$W_{ii} = \frac{1}{\sigma_i} \quad (i = 1,2...\mathrm{m}) \tag{6.11}$$

In general, if a measurement i has a small standard deviation then W_{ii} becomes large and thus this measurement has a relatively large contribution the error function $S(\mathbf{x})_w$. Conversely, an inaccurate measurement (high standard deviation) will tend to have smaller contribution into the $S(\mathbf{x})_w$.

The minimization procedure for the function $S(\mathbf{x})_w$ is identical to that for the least-squares problem and yields a similar equation to 6.6.

$$\mathbf{dx} = (\mathbf{WJ}^T\,\mathbf{WJ})^{-1}\,(\mathbf{WJ})^T\,\mathbf{Wdy} \tag{6.12}$$

Logarithmic weighting

In resistivity inversion it is quite common to apply logarithmic weighting. The reason for this is twofold: **a)** logarithmic weighting constrains the subsurface resistivity to be positive, hence avoiding unrealistic solutions involving negative resistivity. (Sasaki, 1982), **b)** subsurface resistivities can vary over several orders of magnitude and a logarithmic weighting can accelerate the convergence of an iterative algorithm (Park and Van, 1991).

To obtain the least-squares solution with logarithmic weighting requires the minimization of the cost function $\tilde{S}(\mathbf{x}) = \tilde{\mathbf{dy}}^T\,\tilde{\mathbf{dy}}$ where $\tilde{\mathbf{dy}}$ is given by

$$\tilde{\mathbf{dy}} = \ln\ \mathbf{y} - \ln\ \mathbf{f}(\mathbf{x}) \tag{6.13}$$

The parameter vector $\mathbf{x}$ is changed to $\mathbf{p}$ such that $p_i = \exp(x_i)$ and i=1,2..n. The minimization results into a relation which has a form similar to equation 6.6

$$\mathbf{dp} = (\tilde{\mathbf{J}}^T\,\tilde{\mathbf{J}})^{-1}\tilde{\mathbf{J}}^T\,\tilde{\mathbf{dy}} \tag{6.14}$$

where $\tilde{\mathbf{J}}$ is a m x n matrix which has elements $\tilde{J}_{ij} = \partial \ln y_i/\partial p_j$. The matrix $\tilde{\mathbf{J}}$ can be readily expressed as a function of the Jacobian matrix $\mathbf{J}$:

$$\tilde{J}_{ij} = \frac{\partial \ln y_i}{\partial p_j} = \frac{\partial \ln y_i}{\partial \ln x_j} = \frac{\partial y_i}{\partial x_j}\frac{x_j}{y_i} = J_{ij}\frac{x_j}{y_i} \tag{6.15}$$

Equation 6.14 can be used to established an iterative procedure similar to the non-linear least-squares. The new resistivity estimate at the k_{th} iteration is given by : $\mathbf{x}_{k+1} = \mathbf{x}_k + \ln \mathbf{dp}_k$.

Combined weighting

Error and logarithmic weighting can be combined. The cost function becomes

$$\tilde{S}(\mathbf{x})_w = (\mathbf{W}_{\ln}\tilde{\mathbf{d}}\mathbf{y})^T \mathbf{W}_{\ln}\tilde{\mathbf{d}}\mathbf{y} \tag{6.16}$$

where now $\mathbf{W}_{\ln}$ is a diagonal matrix whose elements are the variances of the logarithmic observations $\ln y$. The minimization yields

$$\mathbf{dp} = [(\mathbf{W}_{\ln}\tilde{\mathbf{J}})^T (\mathbf{W}_{\ln}\tilde{\mathbf{J}})]^{-1}(\mathbf{W}_{\ln}\tilde{\mathbf{J}})^T \mathbf{W}_{\ln}\tilde{\mathbf{d}}\mathbf{y} \tag{6.17}$$

6.1.3 Singular value decomposition (SVD)

It has already been explained that the ill-conditioning is due to linear combinations of the unknown parameters which are weakly represented in the measurement system. Under the presence of data noise, these weakly represented linear combinations of parameters not only fail to convey any useful information, but also "contaminate" the entire solution.

One way to avoid this problem is to identify these combinations (by getting a numerical estimate of the system's eigenvalues) and subsequently ignore them. This can be achieved by the use of the SVD technique. The detailed theoretical development of the SVD is described in many texts (Lanczos, 1960; Lawson and Hanson, 1974) and so only an outline of the technique is presented here.

Assume a linear system $\mathbf{A}\mathbf{x}_1 = \mathbf{y}_1$ where $\mathbf{A}$ is a m x n matrix and the adjoint problem $\mathbf{A}^T\mathbf{x}_2 = \mathbf{y}_2$. Those two systems can be combined into a single system

$$\begin{bmatrix} 0 & \mathbf{A} \\ \mathbf{A}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x}_2 \\ \mathbf{x}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \quad \text{or}$$

$$\mathbf{S}\,\mathbf{x} = \mathbf{y} \tag{6.18}$$

Where $\mathbf{S}$ is a k x k (k=m+n) matrix. The general eigenvalue problem for the system of equation 6.18 is $\mathbf{S}\mathbf{w} = \lambda\mathbf{w}$ where $\mathbf{w}$ is an eigenvector and λ an eigenvalue. By

appropriately partitioning $\mathbf{w}=(\mathbf{u},\mathbf{v})$ the following set of equations is obtained

$$\mathbf{A}\ \mathbf{v} = \lambda\mathbf{u} \tag{6.19}$$

$$\mathbf{A}^T\ \mathbf{u} = \lambda\mathbf{v} \tag{6.20}$$

Lanczos (1960) showed that the problem has n+m eigenvectors belonging to n+m independent eigenvalues. He also showed that these are m independent vectors $\mathbf{u}_i$, and n independent vectors $\mathbf{v}_i$. By arranging the normalized eigenvectors in successive columns two square matrices can be obtained: a m x m matrix $\mathbf{U}=(\mathbf{u_1}, \mathbf{u_2}, \ldots \mathbf{u_m})$ which is associated with the (measurement) M-space and a n x n $\mathbf{V}=(\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v_n})$ associated with the (parameter) N-space. The eigenvalue problem described by equations 6.19,6.20 can be written:

$$\mathbf{A}\,\mathbf{V} = \mathbf{U}\mathbf{\Lambda} \tag{6.21}$$

$$\mathbf{A}^T\,\mathbf{U} = \mathbf{V}\mathbf{\Lambda}^T \tag{6.22}$$

where $\mathbf{\Lambda}$ is a n x n diagonal matrix consisting of eigenvalues. Note that the matrices U,V consist of eigenvectors and hence are orthogonal: $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$. Post-multiplying equation 6.21 by $\mathbf{V}^T$ yields

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \tag{6.23}$$

Equation 6.23 expresses the SVD principal [5]. The shape of the matrices involved in the SVD can be seen in Figure 6.2.

As long as the SVD has been calculated, the inverse of the matrix $\mathbf{A}$ can be readily found a (the matrices $\mathbf{U}$, $\mathbf{V}$ are orthogonal and $\mathbf{\Lambda}$ is diagonal):

$$\mathbf{A}^{-1} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}^T \tag{6.24}$$

---

[5] In the more general case where the decomposition produces some zero eigenvalues (let p be the number of non-zero eigenvalues) the decomposition becomes (Lines and Treitel, 1984): $\mathbf{A} = \mathbf{U}_p\mathbf{\Lambda}_p\mathbf{V}_p^T$, where now $\mathbf{U}_p, \mathbf{V}_p$ are semi-orthogonal matrices with dimensions m x p, n x p respectively, and $\mathbf{\Lambda}_p$ is a p x p diagonal matrix. Note that for the resistivity case (as long as a reasonable parametrization scheme is followed) there are no zero eigenvalues.

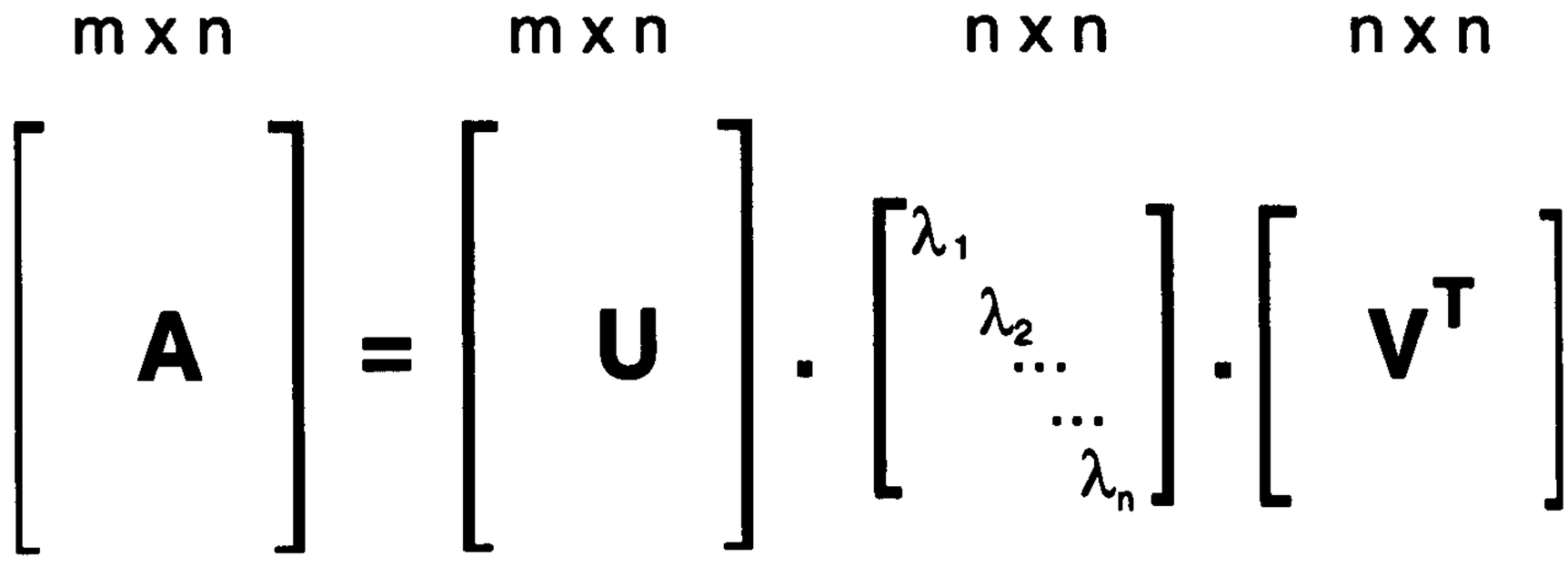


Figure 6.2: The general shape of the matrices involved in the SVD calculation (after Press et al., 1987).

where $\mathbf{\Lambda}^{-1}$ is given by

$$\begin{bmatrix} \frac{1}{\lambda_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\lambda_2} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & & \frac{1}{\lambda_n} \end{bmatrix} \tag{6.25}$$

The solution of the non-linear least-squares problem can be expressed in terms of the SVD formulation. Due to equations 6.23, 6.24 the least-squares correction vector at every iteration is given by (see equation 6.7)

$$\begin{aligned} \mathbf{dx} &= (\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T\mathbf{dy} \\ &= (\mathbf{V}\mathbf{\Lambda}^2\mathbf{V}^T)^{-1}\mathbf{V}\mathbf{\Lambda}\mathbf{U}^T\mathbf{dy} \\ &= \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}^T\mathbf{dy} = \mathbf{J}^{-1}\mathbf{dy} \end{aligned} \tag{6.26}$$

Following Lines and Treitel (1984) the solution vector can be expressed as a weighted vector product sum:

$$\begin{aligned} \mathbf{dx} &= \mathbf{v}_1\frac{1}{\lambda_1}\mathbf{u}_1^T\mathbf{dy} + \mathbf{v}_2\frac{1}{\lambda_2}\mathbf{u}_2^T\mathbf{dy} + \ldots + \mathbf{v}_n\frac{1}{\lambda_n}\mathbf{u}_n^T\mathbf{dy} \\ &= \mathbf{v}_1\frac{a_1}{\lambda_1} + \mathbf{v}_2\frac{a_2}{\lambda_2} + \ldots + \mathbf{v}_n\frac{a_n}{\lambda_n} \end{aligned} \tag{6.27}$$

where $a_i = \mathbf{u}_i^T\mathbf{dy}$ (i=1,2 ..n). From equation 6.27 it is obvious that in general a small eigenvalue λ_i will cause the factor a_i/λ_i to be large so that this term will have a large influence on the solution vector. Due to the limited accuracy of the term a_i (a_i can be inaccurate since it is associated with the measurement vector $\mathbf{y}$) the effect of the factor a_i/λ_i can hence be magnified and "contaminate" the solution.

Solving the least-squares problem using the SVD technique (as described by equation 6.26) will produce solutions which are as unstable as those produced by any matrix inversion technique. The difference lies to that SVD gives all the information needed to handle the problem: by checking the matrix $\mathbf{\Lambda}$ the near-singular eigenvalues can be identified and subsequently excluded from the solution vector (equation 6.27). This is done by setting a threshold value λ_t:

$$\text{if } \lambda_i \leq \lambda_t \text{ then } \frac{1}{\lambda_i} = 0 \quad (i = 1, 2, ..n)$$

In this way linear combinations of parameters which are weakly represented into the system are omitted and the solution becomes stable. The effectiveness of the procedure depends on the successful choice of the threshold: a high threshold results into discarding information which is potentially useful while conversely, a low threshold can produce instability.

Limited use of this procedure has been reported in the solution of the inverse resistivity problem (Inman et al, 1973; Murai and Kagawa, 1985). The reason is that less "drastic" techniques, which do not involve the direct exclusion of the small eigenvalues are preferred. But, as will be shown in the following section, the SVD technique can be used with these techniques as well.

6.1.4 Damped least-squares (Marquadt-Levenberg) method

One way for treating ill-conditioning is to introduce constraints into the optimization procedure instead of directly rejecting the "small" eigenvalues. Treatment of non-linear problems with such an approach was proposed by Marquadt (1963). The technique is known as the damped least squares method (it is also known as the Levenberg-Marquadt method, the Marquadt method, or the Ridge-regression technique) since by constraining the optimization, unbounded oscillations in the solution vector are avoided. The technique has been widely used in the solution of optimization problems which involve ill-conditioned matrices (Golub and Van Loan, 1989).

In the non-linear case the damped least-squares technique involves the following minimization (Box and Camenasu, 1972): minimize the sum of square errors $S(\mathbf{x}) = \mathbf{dy}^T\,\mathbf{dy}$ under the constraint that the amplitude of the solution vector $\mathbf{dx}$ is bounded by a defined quantity, say x_b so that $\mathbf{dx}^T\,\mathbf{dx} \leq x_b$. This type of constrained minimization can be solved by the technique of Lagranian multipliers [6]. The function to be minimized becomes:

$$K(\mu) = \mathbf{dy}^T\mathbf{dy} + \mu(\mathbf{dx}^\mathbf{T}\mathbf{dx} - x_b) \tag{6.28}$$

where μ is the Lagranian multiplier (widely known as the Marquadt's damping factor). The minimization is achieved by setting the derivative of $K(\mu)$ with respect of $\mathbf{dx}$ to zero. This yields (recall equations 6.3, 6.4, 6.5):

$$\begin{aligned} \mathbf{J}^T\,\mathbf{J}\mathbf{dx} + \mu\mathbf{dx} &= \mathbf{J}^T\mathbf{dy} \quad \text{or} \\ (\mathbf{J}^T\,\mathbf{J} + \mu\mathbf{I})\mathbf{dx} &= \mathbf{J}^T\mathbf{dy} \end{aligned} \tag{6.29}$$

where $\mathbf{I}$ is the identity matrix. The solution vector $\mathbf{dx}$ is given by:

$$\mathbf{dx} = (\mathbf{J}^T\,\mathbf{J} + \mu\mathbf{I})^{-1}\mathbf{J}^T\mathbf{dy} \tag{6.30}$$

Marquadt (1963) demonstrated the hybridic nature of the damped least-squares technique. If the Lagranian multiplier is very small ($\mu \rightarrow 0$) then equation 6.30 becomes equivalent to the Gauss-Newton technique (see equation 6.7). If μ is very large ($\mu \rightarrow \infty$) then equation 6.30 becomes equivalent to $\mathbf{dx} \approx \mu^{-1}\mathbf{J^T dy}$. This is proportional to the steepest descent vector [7] which is given by: $\mathbf{dx} = 2\mathbf{J^T dy}$

Box and Kanemasu (1972) discussed the geometrical representation of the damped least-squares method for a two parameter space (p_1, p_2). In such a case the least-

[6] The technique of the Lagranian multipliers (Smith, 1974) solves the problem of minimization of a function under certain restricting conditions by adding these conditions, multiplied by a factor μ, to the initial function and then proceeding with the minimization as if no constraint exists.

[7] The steepest descent method (Press et al. 1987) involves the iterative minimization of a cost function (i.e. the sum square of errors). This is achieved by setting a starting point and finding the direction of the maximum negative gradient of the cost function and then stepping in this direction until a point is reached at which the gradient is minimum. This point will be the new starting point and the entire procedure is repeated until optimization is achieved.

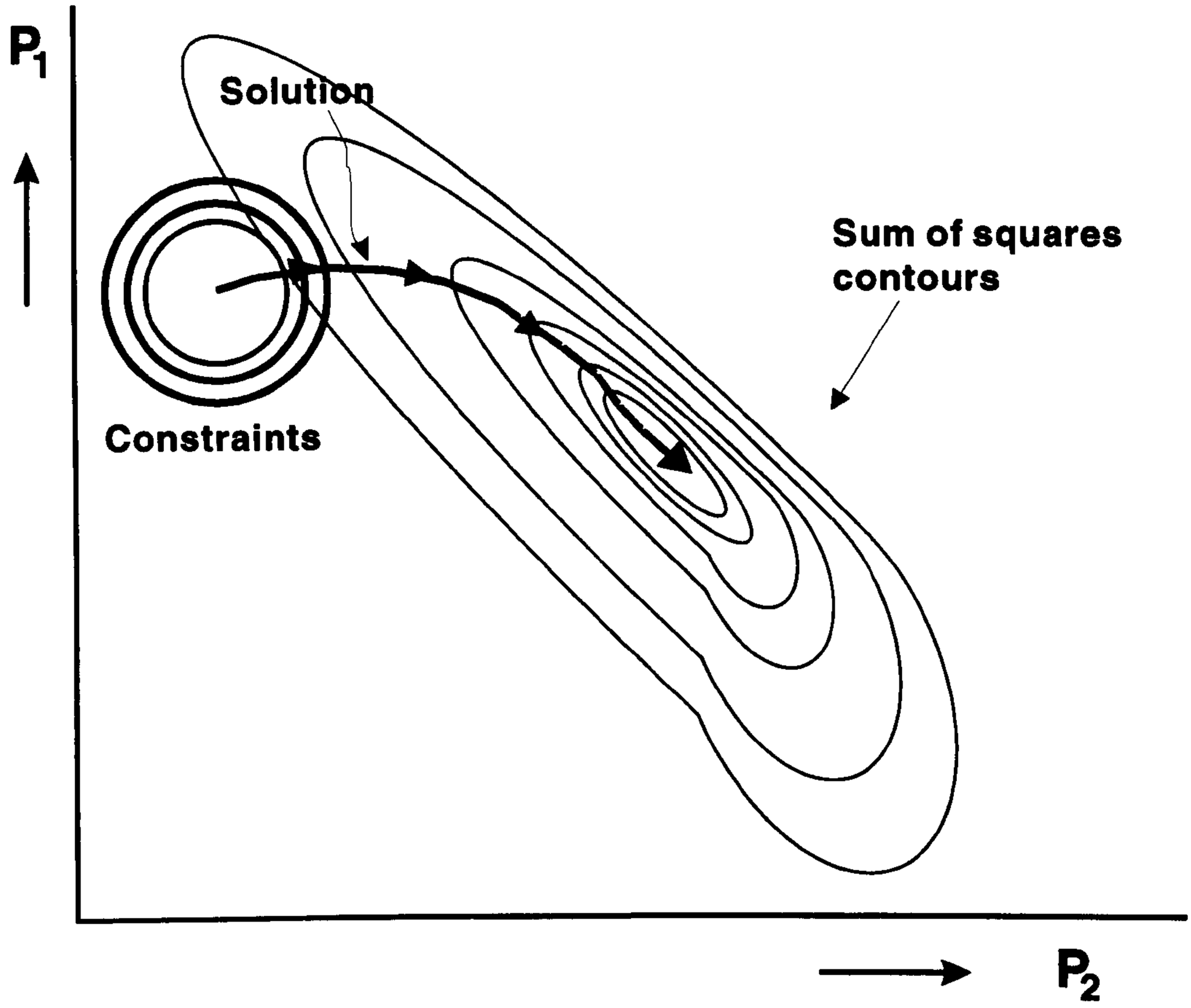


Figure 6.3: Two parameter space representation of damped least-squares minimization (after Box and Kanemasu, 1972).

squares sum is presented as a set of deformed elliptical contours and the damped least-squares solution amounts to minimizing the sum of squares under a circular constraint whose axes are parallel to the p_1, p_2 axes. This can be seen in Figure 6.3.

The only difference between equation 6.30 and the least-squares solution (equation 6.6) is that a constant μ is added to the diagonal elements of the $\mathbf{J}^T\,\mathbf{J}$ matrix. The effect that this addition has on the solution vector can be more clearly demonstrated by the use of SVD.

SVD and damped least-squares

The SVD can be easily applied in combination with the damped least-squares method. Due to equation 6.23, 6.30 becomes

$$\begin{aligned}
\mathbf{dx} &= (\mathbf{V\Lambda U}^T\mathbf{U\Lambda V}^T + \mu\mathbf{I})^{-1}\mathbf{V\Lambda U}^T\mathbf{dy} \\
&= (\mathbf{V\Lambda}^2\mathbf{V}^T + \mu\mathbf{I})^{-1}\mathbf{V\Lambda U}^T\mathbf{dy} \\
&= \mathbf{V}(\mathbf{\Lambda}^2 + \mu\mathbf{I})^{-1}\mathbf{V}^T\mathbf{V\Lambda U}^T\mathbf{dy} \\
&= \mathbf{V}\mathrm{diag}\left(\frac{1}{\lambda_i^2+\mu}\right)\mathbf{\Lambda U}^T\mathbf{dy} \\
&= \mathbf{V}\mathrm{diag}\left(\frac{\lambda_i}{\lambda_i^2+\mu}\right)\mathbf{U}^T\mathbf{dx} \qquad (6.31)
\end{aligned}$$

where $\mathrm{diag}[\lambda_i/(\lambda_i^2+\mu)]$ is given by:

$$\begin{bmatrix}
\frac{\lambda_1}{\lambda_1^2+\mu} & 0 & \dots & 0 \\
0 & \frac{\lambda_2}{\lambda_2^2+\mu} & & \vdots \\
\vdots & & \ddots & \vdots \\
0 & \dots & & \frac{\lambda_n}{\lambda_n^2+\mu}
\end{bmatrix} \qquad (6.32)$$

Equation 6.31 can be written as a weighted vector product sum (recall equation 6.27)

$$\mathbf{dx} = \mathbf{v}_1\frac{\lambda_1}{\lambda_1^2+\mu}\mathbf{u}_1^T\mathbf{dy} + \mathbf{v}_2\frac{\lambda_2}{\lambda_2^2+\mu}\mathbf{u}_2^T\mathbf{dy} + \ldots + \mathbf{v}_n\frac{\lambda_n}{\lambda_n^2+\mu}\mathbf{u}_n^T\mathbf{dy} \qquad (6.33)$$

Equation 6.33 describes the damped least-squares mechanism. If the eigenvalue λ_i (i=1,2..n) is large then the addition of the Lagranian multiplier will have no significant effect on the factor $[\lambda_i/(\lambda_i^2+\mu)]$, and thus no significant effect to the solution vector **dx**. But, if λ_i is small ($\lambda_i \rightarrow 0$) then the addition of the Lagranian multiplier will cause the factor $[\lambda_i/(\lambda_i^2+\mu)]$ to become smaller and thus force it to contribute less in the solution vector **dx**.

Marquadt's algorithm

Marquadt (1963) proposed an iterative damped least-squares scheme for solving non-linear problems. The scheme is quite similar to the iterative least-squares technique

and can be easily adapted to the resistivity problem. Given an initial resistivity estimate $\mathbf{x}_o$ calculate the sum of square errors $S(\mathbf{x}_o)$ and set an initial value of μ_o (say 0.1).

1. At the k_{th} iteration, if $\mathbf{x}_k$ is the resistivity distribution, calculate the forward model response $\mathbf{f}(\mathbf{x}_k)$ and the Jacobian matrix $\mathbf{J}_k$ which corresponds to the $\mathbf{x}_k$ resistivity distribution.

2. Find the parameter correction vector $\mathbf{dx}_k$ which is given by (see equation 6.30)

$$\mathbf{dx}_k = (\mathbf{J}_k^T \mathbf{J}_k + \mu_k \mathbf{I})^{-1} \mathbf{J}_k^T \mathbf{dy}_k \tag{6.34}$$

   where $\mathbf{dy}_k = \mathbf{y} - \mathbf{f}(\mathbf{x}_k)$.

3. Calculate the new sum of square errors $S(\mathbf{x}_k + \mathbf{dx}_k)$

4. If $S(\mathbf{x}_k + \mathbf{dx}_k) \geq S\mathbf{x_k}$ (divergence) then set $\mu_k = \mu_k\, c$ where c is a set factor ($c > 1$, say 10) and go back to step 2.

5. If $S(\mathbf{x}_k + \mathbf{dx}_k) \leq S(\mathbf{x_k})$ (convergence) then set $\mu_{k+1} = \mu_k / c$, update the resistivity estimate $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{dx}_k$ and start a new iteration.

6. The iterative procedure is repeated until a stopping criterion is met (i.e. no significant decrease in error is achieved with further iterations).

If a data-error and logarithmic weighted scheme is used in combination (recall section 6.1.2) then the correction vector of equation 6.35 becomes $\mathbf{dx}$=ln $\mathbf{dp}$ and $\mathbf{dp}$ is given by:

$$\mathbf{dp} = [(\mathbf{W}_{\ln}\tilde{\mathbf{J}})^T (\mathbf{W}_{\ln}\tilde{\mathbf{J}}) + \mu \mathbf{I}]^{-1} (\mathbf{W}_{\ln}\tilde{\mathbf{J}})^T \mathbf{W}_{\ln}\tilde{\mathbf{dy}} \tag{6.35}$$

The quantities involved in equation 6.35 are fully explained in section 6.1.2.

The algorithm exploits the above mentioned hybridic nature of the method: the technique initially is closer to the steepest descent method, which is generally better when the error S($\mathbf{x}$) is large, but as the Lagranian multiplier decreases, it becomes

closer to the Gauss-Newton method, which is more effective when the error S is small (Lines and Treitel, 1984).

Marquadt's algorithm has been used extensively for the solution of the 2-D inverse resistivity problem (Rijo et al, 1977; Pelton et al. 1978; Petrick et al., 1981; Sasaki 1982; Smith and Vozoff 1984; Trip et al., 1984) mainly due to its stability and its fast convergence (quadratic). Jupp and Vozoff (1975) suggested a slightly modified algorithm which applies the Marquadt's technique in combination with the truncation method. Two main disadvantages of the algorithm have been reported: **a)** the final result is strongly dependant on the initial model choice $\mathbf{x}_o$ (Smith and Vozoff, 1984) and **b)** in some cases the technique produces unnecessarily complex solutions - unnecessary in the sense that although they are mathematically correct, they are not physically reasonable.

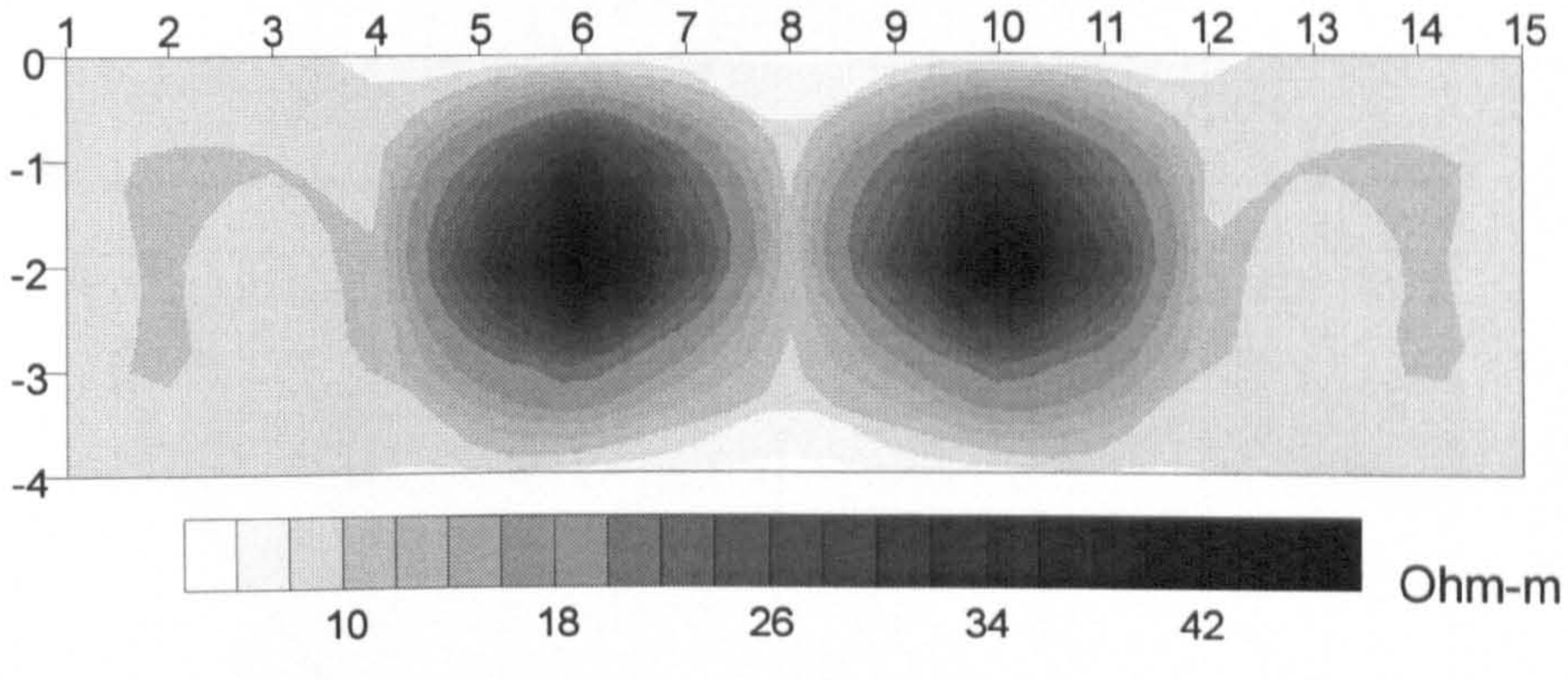
Inversion examples

Here, Marquadt's non-linear optimization algorithm has been applied to a series of synthetic data. The forward calculations were performed by the FEM model and the Jacobian matrix was calculated by the use of the adjoint equation technique (see section 4.3.2). A logarithmic and data-error (when real or noise-corrupted synthetic data was used) weighting was applied (see equation 6.35). In order to avoid forming the terms $\tilde{\mathbf{J}}^T\tilde{\mathbf{J}}$ and $\tilde{\mathbf{J}}^T\mathbf{d}\tilde{\mathbf{y}}$ [8] inferred by equation 6.15, the damped least-squares solution at every iteration was obtained by the use of the SVD (see equation 6.31). Computationally speaking, the SVD procedure is quite complicated. Efficient SVD routines have been presented by Lawson and Hanson (1974), Press et al. (1987); the routine of the latter is used in this work.

In Figure 6.4a the reconstruction of the noise-free dipole-dipole data set which corresponds to Model 1 is depicted. The quality of the reconstruction is particularly good

[8]Lines and Treitel (1984) argued that the formation of those terms involves inaccuracies which could be problematic when large Jacobian matrices are considered.

(a)



(b)

1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	5.3	5.1	1.0	1.0	5.1	5.3	1.0	1.0	1.0	1.0
1.0	9	9	1.4	3.4	3.0	1.3	1.3	3.0	3.4	1.4	9	9	1.0
1.0	1.0	9	1.0	1.1	1.1	9	9	1.1	1.1	1.0	9	1.0	1.0

Figure 6.4: Inversion results of Model 1 (dipole-dipole data, 5 iterations) using the Marquadt's method: a) in a gray scale form, b) in a post map form.

- the resulting resistivities are quite close to the initial model (see Figures 6.4b and 6.1b). The convergence pattern for this inversion is depicted in Figure 6.5. Similarly good results were obtained for the inversion of Model 2 (Figure 6.6a,b) and Model 3 (Figure 6.6c,d). All inversions used a homogeneous ground as a starting model (the average of the measured apparent resistivities was the homogeneous ground resistivity) and produced an extremely low % RMS error (less that 0.5%). Note that the parameters at the bottom edges are somewhat less well recovered, since they are not so well resolved (comparatively irrelevant parameters). When noise-corrupted data were inverted (5% additive Gaussian noise) Marquadt's method did not perform so well. As can be seen in Figure 6.7a,b although the inversion of Model 1 still delineates the targets successfully there are also spurious noise-related artifacts in the final reconstruction. Similar artifacts were observed in all test cases. Note also that in those cases the algorithm became less stable: several changes in the Marquadt factor had to be made in order to achieve a better reconstruction.

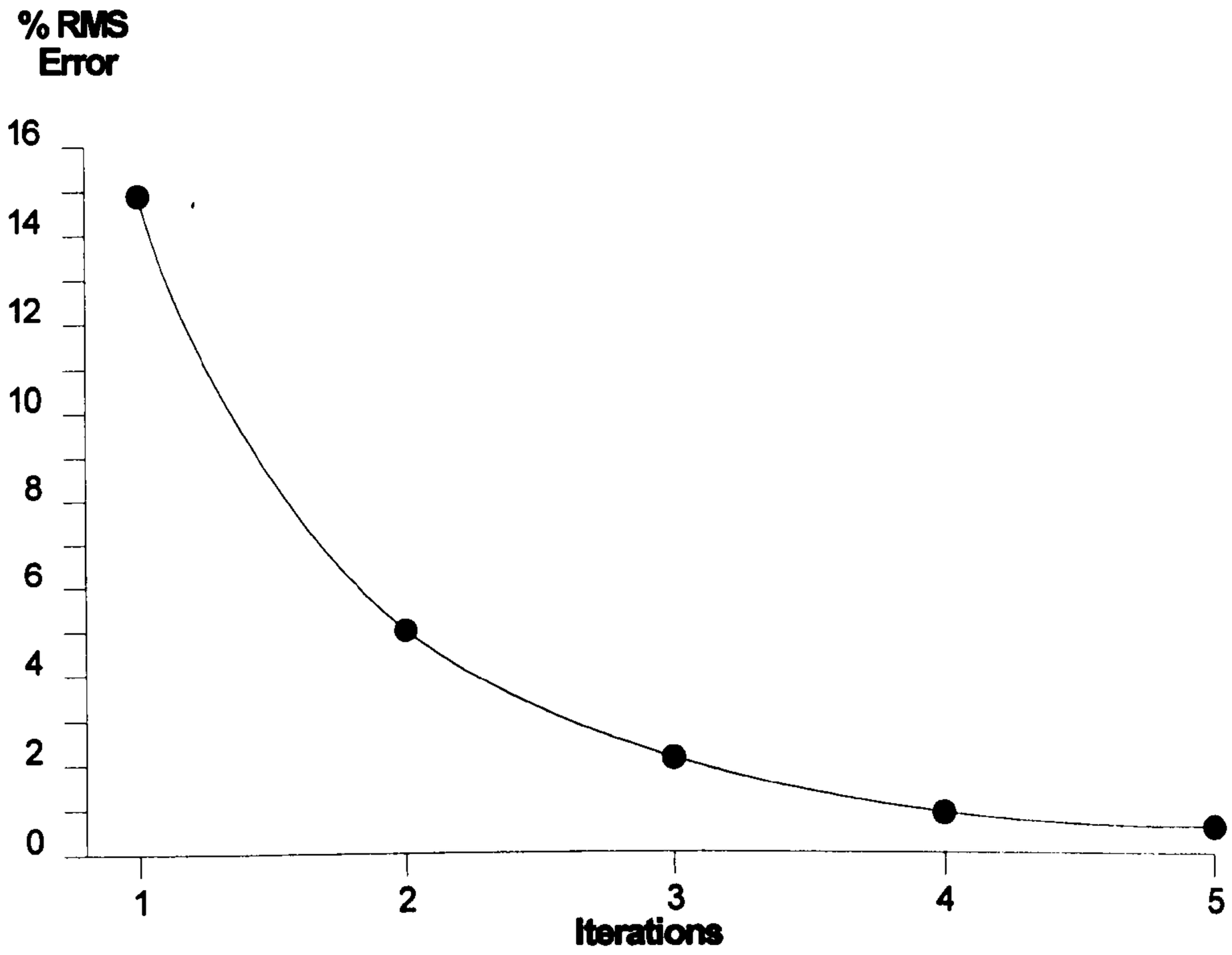


Figure 6.5: The variation of the RMS error with iterations for the Marquadt's inversion of Model 1.

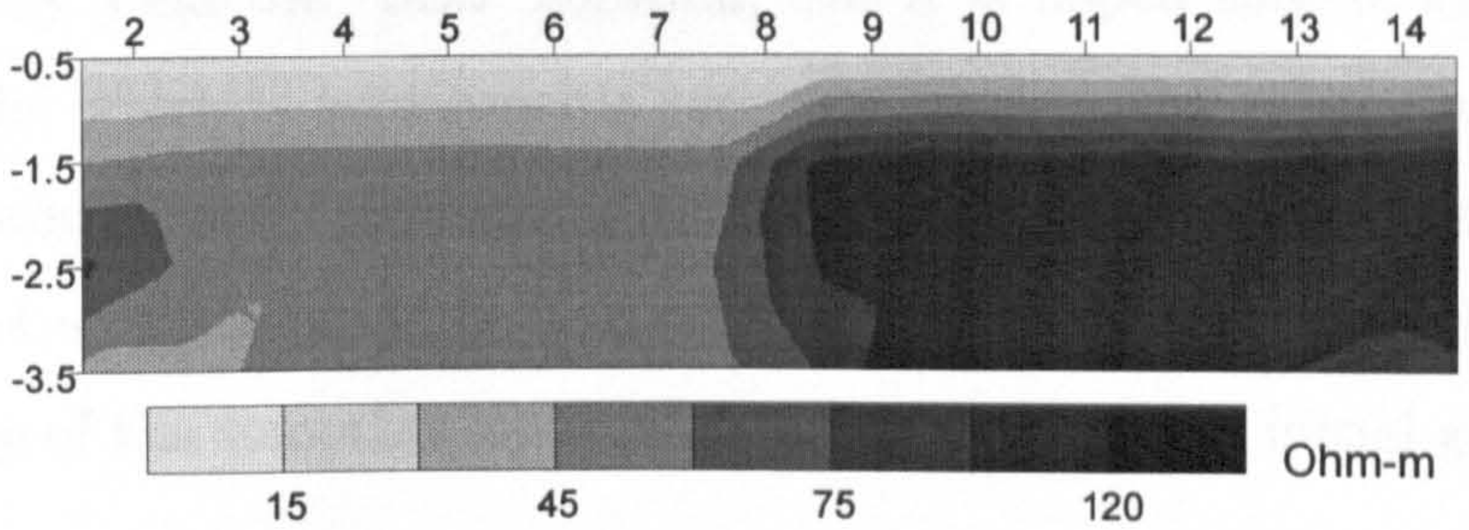
6.1.5 Smoothness constrained (Occam) inversion

Another way to tackle the instability of the resistivity inverse problem is to impose a smoothness constraint. The use of such a constraint to stabilize poorly conditioned problems belongs to a general class of methods known as regularization techniques (Tikhonov, 1963) [9]. The technique has been proposed for geophysical purposes by Constable et al. (1987) who called it Occam's inversion (after the 14th century philosopher) and applied it to the 1-D resistivity and magnetotelluric inverse problems. The idea is to find the smoothest model which could fit the data in the sense that the model should depart from the simplest case only as far as necessary to fit the data.

The need for such a constraint derives from the reported deficiencies of the damped

[9]This type of constrained optimization is also known as the LSQI problem (least-squares minimization with quadratic inequality constraint) (Golub and Van Loan, 1989).

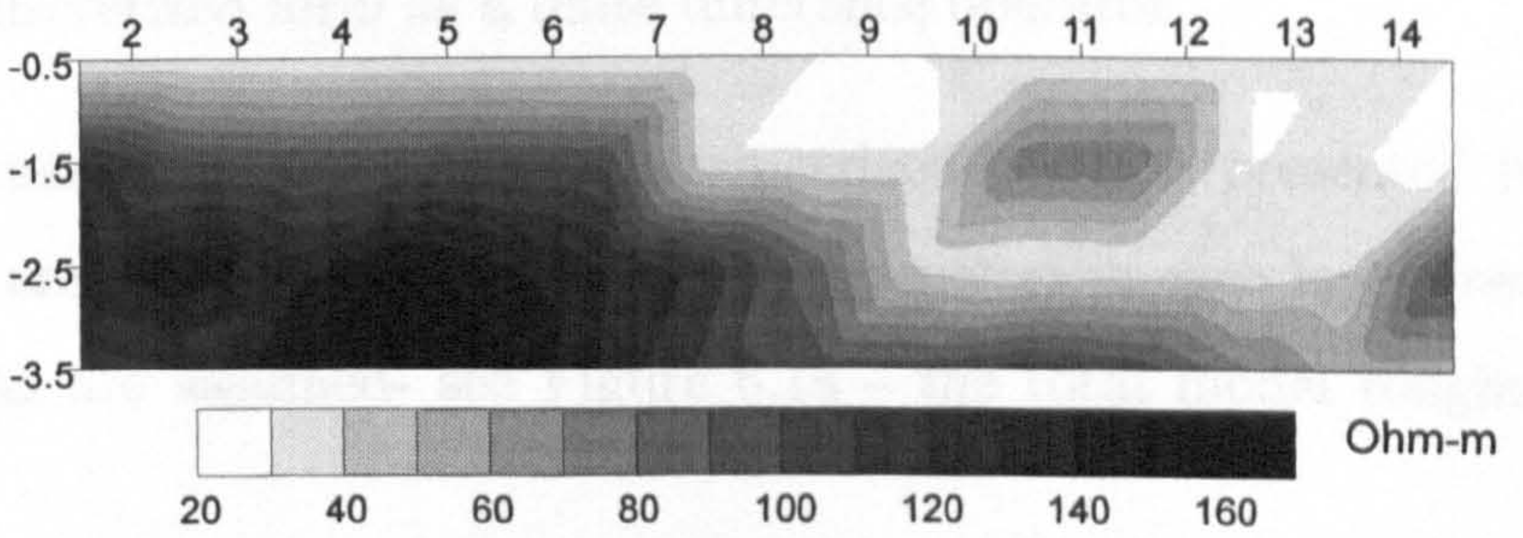
(a)



(b)

1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
4.5	5.0	5.0	5.0	5.0	4.9	4.9	9.5	9.9	9.6	9.9	9.4	9.8	10.6
7.9	5.8	4.6	5.3	4.9	4.7	5.9	9.2	10.0	8.9	9.8	9.7	12.1	10.6
4.1	4.1	4.9	4.9	5.1	5.0	5.4	7.6	11.1	12.2	10.5	9.8	7.8	8.6

(c)



(d)

3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
12.7	10.3	9.9	10.0	10.1	9.9	3.0	3.0	3.0	7.9	8.0	3.0	3.1	2.7
13.7	15.0	13.4	15.3	14.2	13.1	11.6	8.4	3.5	2.8	3.1	3.2	3.7	10.3
11.3	12.0	15.3	16.7	15.2	12.8	10.6	9.5	9.4	10.5	9.7	7.2	5.5	5.4

Figure 6.6: Inversion results using the Marquadt's method. Inversion of Model 2 (dipole-dipole data, 6 iterations) a) in grey scale form, b) in post map form. Inversion results of Model 3 (dipole-dipole data, 6 iterations): c) in a grey scale form, d) in a post map form.

least-squares technique, namely the dependence of the result on the starting guess and the unnecessarily complex solutions. The smoothness constrained inversion will not necessarily yield the "best" solution, but it is hoped that it will produce a simplified model which is a reasonable representation of the subsurface. At the same time, the smoothness constraint guarantees stability and, most importantly, produces a model which is based on a characteristic that the user has chosen (the pattern or nature of the smoothness) and not on some arbitrary initial guess.

The roughness term

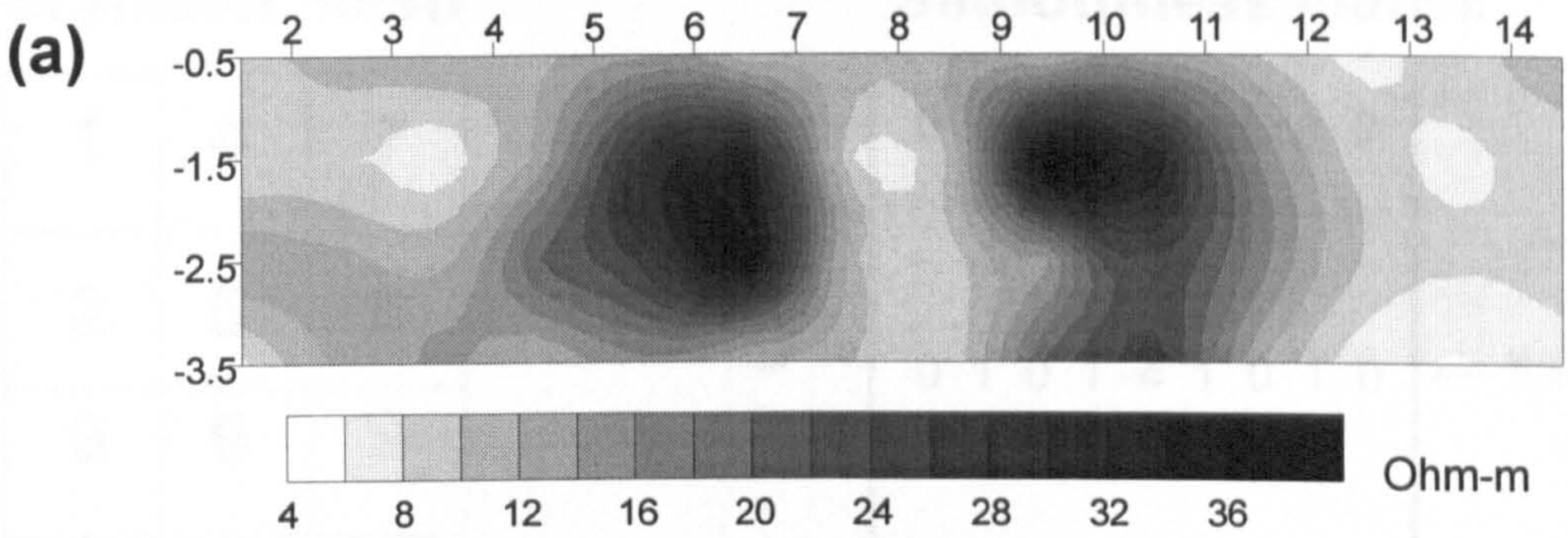
The smoothness constrained minimization can be achieved by firstly defining a term which describes the desired smoothness relationships between the parameters as a function of their resistivity. This roughness term (Constable et al., 1987) can be expressed in a discretized form as a finite difference operator.

A simple roughness term for the 2-D parametrized earth is presented by Sasaki (1992): If there are L parameter layers and Q parameters in each layer (rectangular parameter blocks are assumed- see Figure 6.1a - the total model roughness R is given by:

$$R = \sum_{k=2}^{L-1} \sum_{l=2}^{Q-1} \left[x_{(k-1,l)} + x_{(k,l-1)} - 4x_{(k,l)} + x_{(k,l+1)} + x_{(k+1,l)} \right]^2 \qquad (6.36)$$

Where $x_{(kl)}$ is the resistivity of the parameter which belongs at the k_{th} layer and l_{th} column and $x_{(k-1,l)}$, $x_{(k,l-1)}$, $x_{(k,l+1)}$, $x_{(k+1,l)}$ are the resistivities of the western, northern, southern and eastern immediate neighbours of the $x_{(k,l)}$ parameter respectively [10]. From equation 6.36 it is clear that the larger the resistivity variation among neighbouring parameters the larger the factor R will be.

[10]Note also that in equation 6.36 the roughness is expressed as sum of central differences and therefore as a sum of the second derivatives of the resistivities in respect of the x and z coordinates. deGroot-Hedlin and Constable (1990) proposed a 2-D model expressed as a sum of right-hand side differences - that is, as a sum of first derivatives. From some preliminary tests we concluded that there was no apparent difference between using first or second derivatives as far as the results are concerned.



(b)

9	11	12	12	14	10	11	12	10	12	10	7	9	12
8	8	6	13	28	33	8	9	41	32	16	9	7	9
14	13	11	21	25	37	11	9	14	24	17	10	8	9
7	13	15	11	16	11	10	8	22	26	12	7	6	7

Figure 6.7: Inversion results using the Marquadt's method. Inversion of Model 1 (dipole-dipole data, 5% random noise, 4 iterations, error 5.4%): a) in grey scale form, b) in post map form.

Equation 6.36 can be also written in a matrix form

$$R = (\mathbf{C}\,\mathbf{x})^T (\mathbf{C}\,\mathbf{x}) \tag{6.37}$$

where $\mathbf{C}$ is a n x n banded matrix (n is the number of parameters). The i_{th} row of the matrix contains the smoothness information for the i_{th} parameter and will have all its elements zero except from those corresponding to this parameter (this element will have a value of -4) and its immediate neighbours (those elements will have a value of 1). In Figure6.8 the matrix $\mathbf{C}$ for a simple case of 9 parameters can be seen.

The optimization algorithm

The initial non-linear problem can again be linearized by the use of a Taylor series expansion and treated as a series of linear problems. Since the solution of each

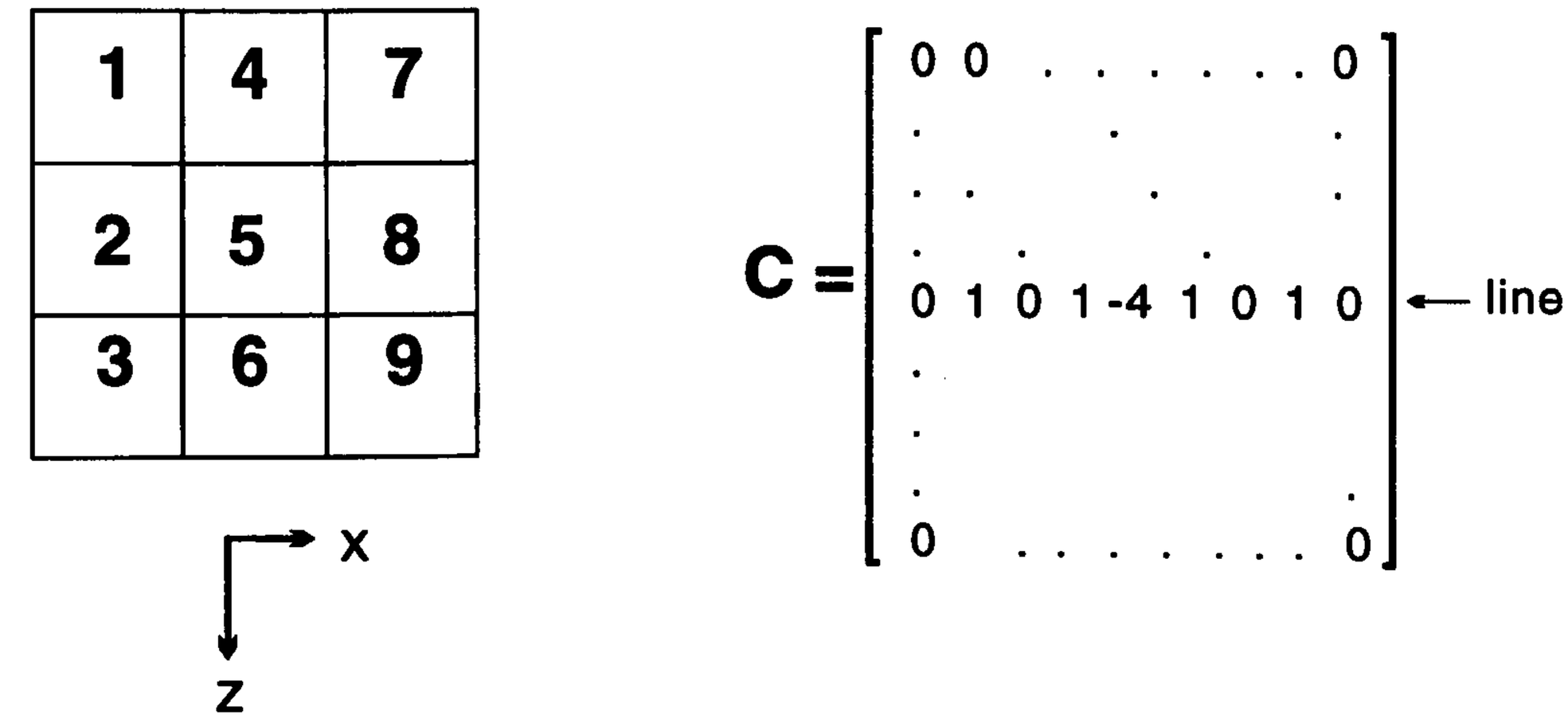


Figure 6.8: The smoothness matrix (right) for the case of a 9 parameter mesh (left).

linear problem will produce a resistivity correction vector **dx** which is the difference between two subsequent models, the roughness term can be defined in respect of the correction vector **dx**. In other words, the smoothness in the final model is achieved by imposing the smoothness in each model correction. The roughness term R (see equation 6.37) is redefined as ΔR such that

$$\Delta R = (\mathbf{C}\,\mathbf{dx})^T\,(\mathbf{C}\,\mathbf{dx}) \tag{6.38}$$

In every iteration the following minimization is sought (Constable et al, 1984): minimize ΔR under the constraint that the sum of square errors S($\mathbf{x}$) is minimized or (since we are dealing with real data) becomes equal to X_d - the acceptable error in view of data uncertainties [11]. In other words, find the least rough resistivity correction vector which will fit the data. By using a Lagranian multiplier μ^{-1} the cost function becomes

$$\begin{aligned} U &= \Delta R + \mu^{-1}(S - X_d) \\ &= (\mathbf{C}\,\mathbf{dx})^T\,(\mathbf{C}\,\mathbf{dx}) + \mu^{-1}\left[(\mathbf{dy} - \mathbf{Jdx})^T(\mathbf{dy} - \mathbf{Jdx}) - X_d\right] \end{aligned} \tag{6.39}$$

[11] Any effort to continue minimization beyond this poind would result in potentially fitting the data to noise.

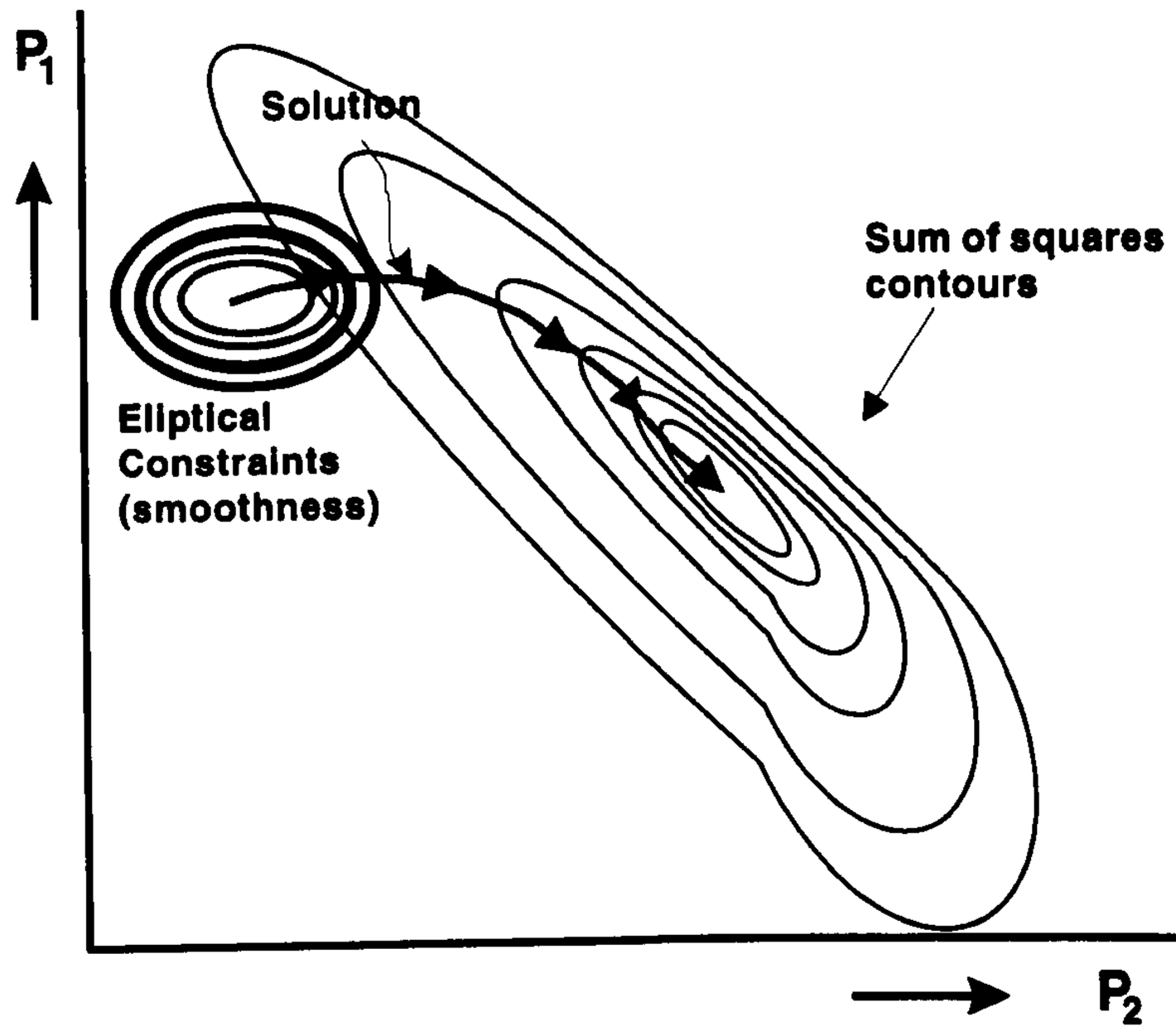


Figure 6.9: Two parameter space representation of constrained minimization (after Box and Kanemasu, 1972).

Setting the derivative of U with respect of **dx** equal to zero and performing the calculations finally yields

$$(\mathbf{J}^T\,\mathbf{J} + \mu\mathbf{C}^T\,\mathbf{C})\mathbf{dx} = \mathbf{J}^T\,\mathbf{dy} \tag{6.40}$$

Note that the only difference in equation 6.39 compared to the damped least-squares technique (see equation 6.33) is that the identity matrix **I** is replaced by the smoothness matrix $\mathbf{C}^T\mathbf{C}$. Box and Kanemasu (1972) suggested that solutions of the form of equation 6.40 are a generalized case of Marquadt's algorithm. In such a case, and for a two parameter space (p_1, p_2), the least-squares sum is presented as a set of deformed elliptical contours and the solution amounts to minimizing the sum of squares under an elliptical constraint whose axes are parallel to the p_1, p_2 axes. This can be seen in Figure 6.9.

An iterative algorithm can now be established as follows:
Produce a matrix **C** which describes the smoothness pattern of the model, set an initial model $\mathbf{x_0}$ and an initial Lagranian multiplier value μ_0.

1. At the k_{th} iteration the parameter correction vector $\mathbf{dx_k}$ is given by

$$\mathbf{dx}_k = (\mathbf{J}_k^T \mathbf{J}_k + \mu_k \mathbf{C}^T \mathbf{C})^{-1} \mathbf{J}_k^T \mathbf{dy}_k \tag{6.41}$$

where $\mathbf{J}_k$ is the Jacobian matrix which corresponds to the $\mathbf{x}_k$ resistivity distribution, and $\mathbf{dy}_k = \mathbf{y} - \mathbf{f}(\mathbf{x}_k)$[12].

2. The new resistivity estimate $\mathbf{x}_{k+1}$ is given by

$$\mathbf{x_{k+1}} = \mathbf{x_k} + \mathbf{dx_k} \tag{6.42}$$

3. The iterative procedure is repeated until a stopping criterion is met (e.g. no significant decrease in error is achieved with further iterations, or the error is just above the estimated observation uncertainties)

When data-errors and logarithmic weighting are considered (recall section 6.1.2) the resistivity correction vector becomes $\mathbf{dx}$=ln $\mathbf{dp}$, and $\mathbf{dp}$ is given by:

$$\mathbf{dp} = [(\mathbf{W}_{\ln}\tilde{\mathbf{J}})^T (\mathbf{W}_{\ln}\tilde{\mathbf{J}}) + \mu \mathbf{C}^T \mathbf{C}]^{-1} (\mathbf{W}_{\ln}\tilde{\mathbf{J}})^T \mathbf{W}_{\ln}\tilde{\mathbf{dy}} \tag{6.43}$$

The Lagranian multiplier (LM) is one way to control the smoothness of the inverted model: the larger the value of the LM the smoother the model, but how to decide the LM value (μ_k)? A reasonable scheme is to start the inversion with a relatively high LM value (increased smoothness) and reduce it gradually at each iteration (decrease the smoothness weighting as the final solution is approached).

Constable et al. (1987) suggested the following technique: at the beginning of every iteration test several values of μ (chosen on a linear scale for example) and calculate

---

[12]Note that if the roughness term was chosen to be related to the resistivity model (equation 6.37) and not to the resistivity correction vector (equation 6.38) the minimization would have produced:

$\mathbf{dx}_k = (\mathbf{J}_k^T \mathbf{J}_k + \mu_k \mathbf{C}^T \mathbf{C})^{-1} \mathbf{J}_k^T (\mathbf{dy}_k + \mathbf{J}_k x_k)$

Xu (1993) showed (via synthetic examples) that this type of minimization produces less satisfactorily results with noisy data.

for each one the resistivity correction vector (using equation 6.41). Then calculate the sum of errors $S(\mathbf{x})$ which correspond to those newly found resistivities. The optimum μ value is chosen as the one that produces the minimum error. A 1-D line search procedure can be used to find the optimum value. This scheme involves the repetition of the inversion procedure (equation 6.41) and of the forward modelling calculations (to find the errors) for as many times as the number of μ values to be tested.

Few applications of the Occam scheme into the 2-D resistivity inversion are presented in the literature. Sasaki (1989) used this scheme to jointly invert resistivity and magentotelluric data. In a later work (1992) he applied the scheme to cross-borehole resistivity data. Xu (1993) compared the effect of different smoothness constraints to the inversion. Application of the technique has been also reported by Daily et al. (1995).

Inversion examples

A 2-D smoothness constrained inversion algorithm has been created in order to allow the 2-D inversion of earth resistivity data. The FEM algorithm and the adjoint equation approach were used again to calculate the forward model and the Jacobian matrix. A simple smoothness pattern (given by equation 6.36) was used.The 1-D line search technique was found to be time consuming. For a simplest line search several μ values have to be tested at every iteration, but this does not guarantee that the optimum value is chosen. We used a scheme that calculated 5 μ values. A cubic spine interpolation algorithm was used to generate the line information and a golden section search was then used to find the optimum μ value. [both interpolation and golden section routines are described in Press et al. (1987)]. In Figure 6.10a,b the resultant spline interpolations for iterations one and two and the optimum μ values chosen for the inversion of Model 1 can be seen. In Figure 6.10c the quadratic convergence of the algorithm for the inversion of Model 1 is shown.

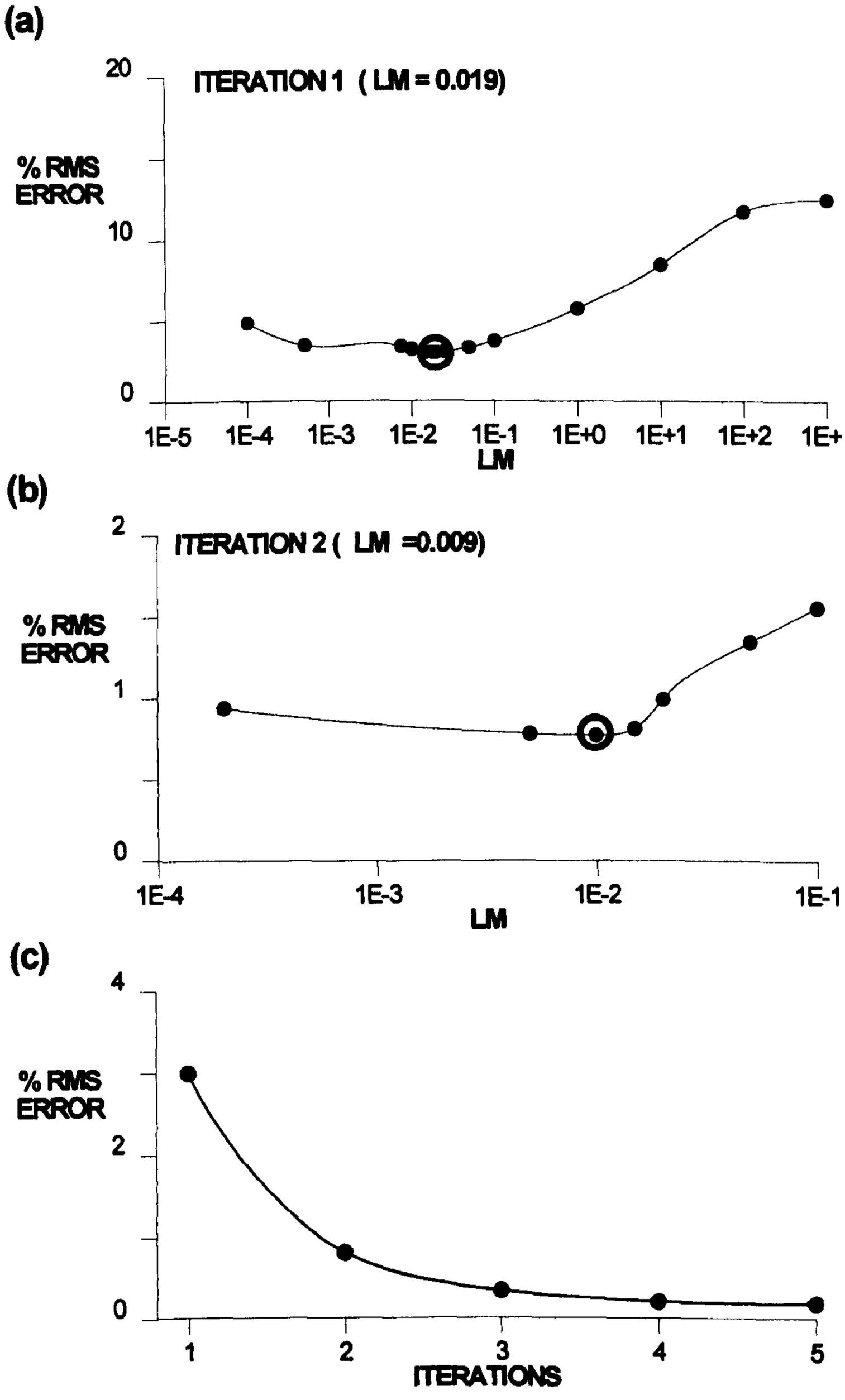


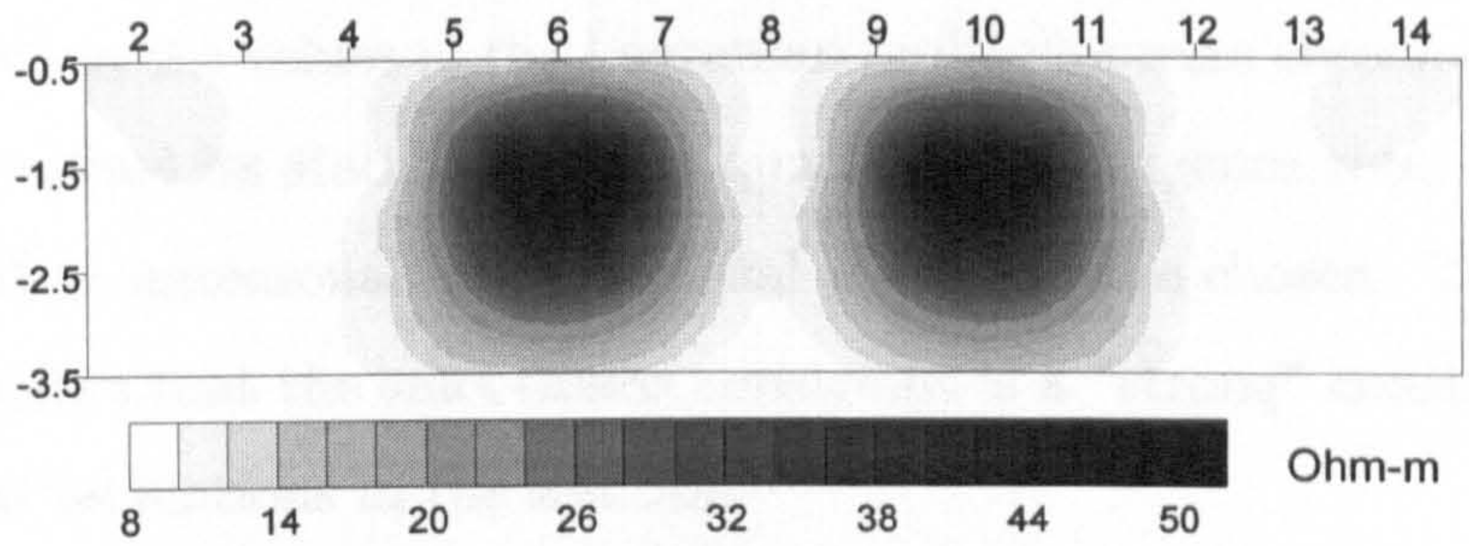
Figure 6.10: The variation of the % RMS error with different Lagranian multiplier values for the inversion of Model 1: a) for iteration 1, b) for iteration 2. c) The variation of the % RMS error with iterations (inversion of Model 1).

The algorithm produced good results when noise-free data were inverted, as it can be seen in the results obtained from the inversion of Model 1 (Figure 6.11a,b) and Model 2 (Figure 6.11c,d). All inversions had a homogeneous ground as a starting model and produced a low % RMS error (less that 0.5%). It is worthwhile noticing the effect that smoothness has in the reconstruction of Model 2: the resistivity varies smoothly between the contrasting media and whilst the resultant image is a reasonable representation of reality, it does not give accurate information about the exact positioning of the media.Due to this effect there might be arguments that smoothness is not always physically reasonable: in cases of abrupt resistivity contrasts, for example, the smoothness constraint may fail to produce the exact boundaries/contrasts of the media. But in 2-D parametrization schemes this is not a problem: the parameters have artificial boundaries chosen by the user, so it is unlikely that the boundaries of the target will coincide with the boundaries of the parameters. Therefore, the boundaries which will show up in the inverted subsurface image will not be exact anyway. In such a case smoothness becomes an advantage since it reflects reality in a safe (from the interpreter's point of view) way.

In Figure 6.12 an example of such a situation is given. The reconstruction of the dipole-dipole data set which corresponds to the resistive prism of Figure 6.12a (note that its lateral boundaries do not coincide with the boundaries of the user-defined parameters) was obtained using both Occam inversion (Figure 6.12b,c) and Marquadt's method (Figure 6.12d,e). The results indicate that Occam inversion produced a realistic image of the structure whilst Marquadt's technique tended to concentrate the resistivity in the upper layer. It is important to note that this result cannot be generalized: Marquadt's technique could have produced different results if a different initial model and/or different Lagranian multiplier values were used. What can be safely said though is that, at least, we would not expect the Occam inversion to produce results which are inferior to Marquadt's technique for this type of situation.

One of the advantages of the smoothness constraint approach is that it is particularly suited to the treatment of noisy data. As can be seen in Figure 6.13a,b, the inversion

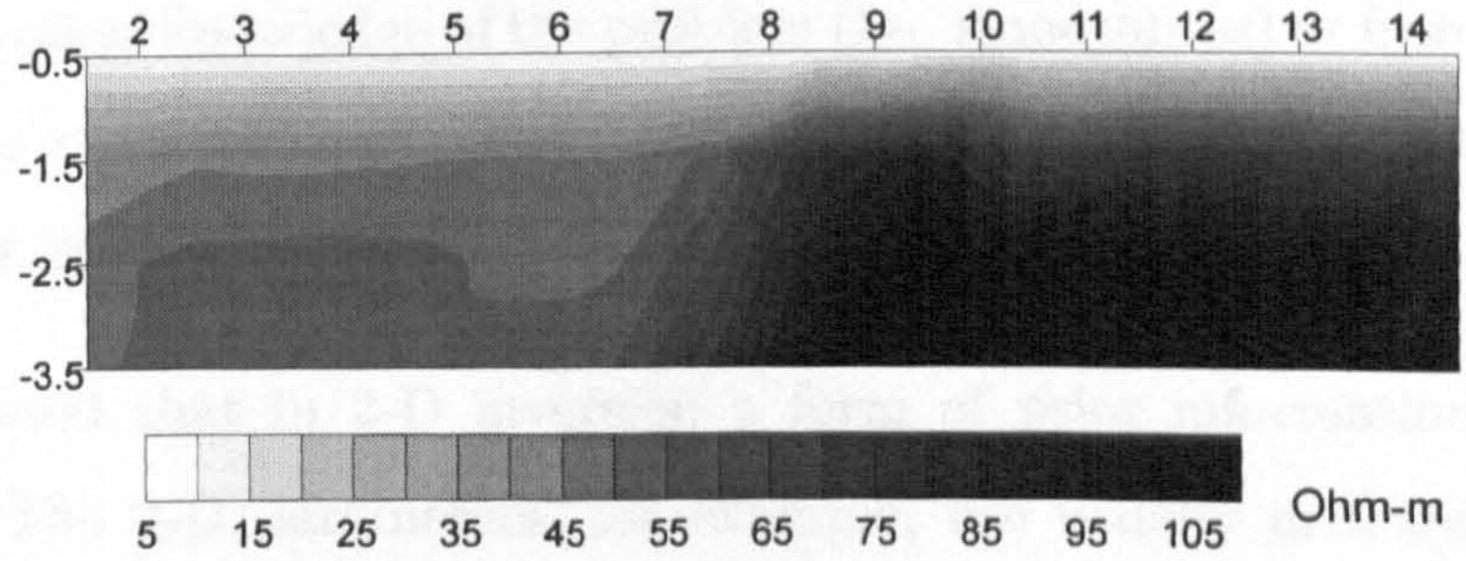
(a)



(b)

1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.1	4.8	5.2	1.1	1.1	5.2	4.8	1.1	1.0	1.0	1.0
1.0	9	1.0	1.3	3.5	3.5	1.2	1.2	3.5	3.5	1.3	9	8	9
1.0	1.0	1.0	1.1	1.3	1.2	9	9	1.2	1.3	1.1	9	9	1.0

(c)



(d)

1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
4.1	4.5	4.4	4.5	4.6	4.5	5.1	7.9	8.9	7.2	7.7	7.6	7.0	7.2
4.8	5.2	5.3	5.2	4.9	4.9	6.1	9.4	10.9	10.8	10.5	10.3	10.4	9.0
4.9	5.4	5.2	5.2	5.2	5.5	7.1	10.3	10.0	10.2	9.8	9.9	9.8	9.9

Figure 6.11: Inversion results of Model 1 (dipole-dipole data, 5 iterations) using the Occam scheme: a) in a grey scale form, b) in a post map form. Inversion results of Model 2 (dipole-dipole data, 6 iterations) using the Occam scheme: a) in a grey scale form, b) in a post map form.

of Model 1 still delineates the targets quite successfully without producing any significant noise-related artifacts. Another impressive feature of the technique is its stability. Even when the values of the Lagranian multiplier were chosen empirically, the technique retained its stability and its quadratic convergence rate. Divergence occurred only when unreasonably small initial μ values were chosen. The stability derives from the fact that the smoothness constraint is a "strong" constraint which damps unwanted oscillations in the solution.

6.1.6 Inversion with *a priori* information

One way to stabilize the inversion procedure is to constrain it in order to be consistent with any existing prior information. The prior information could derive from some general physical knowledge of the problem (i.e. smoothness) or from knowledge of the particular surveyed area. The later could be available from several sources such as different geophysical techniques, drilling, excavation results.

It has to be noted that in 2-D inversion a form of prior information has to be used anyway. The 2-D parameters, for example, are usually arranged in a way which is consistent with the local geology (stratigraphy) and with the depth/size of the targets under investigation. Moreover, for the damped least-squares inversion a good initial guess - which is a result of prior information - is crucial for the successful implementation of the technique. But the problem with such an implementation of prior information in an iterative procedure is that there is no guarantee that the final result will be consistent with the prior information.

The idea behind the inversion with *a priori* information is the following: among the many possible models that could fit the data (there are many models due to the ill-posed nature of the inverse problem and the data uncertainties) choose the one which is as close as possible to the prior information. The use of prior information will stabilize the inverse procedure and as long as the prior information is reasonably accurate will produce a model which is close to reality.

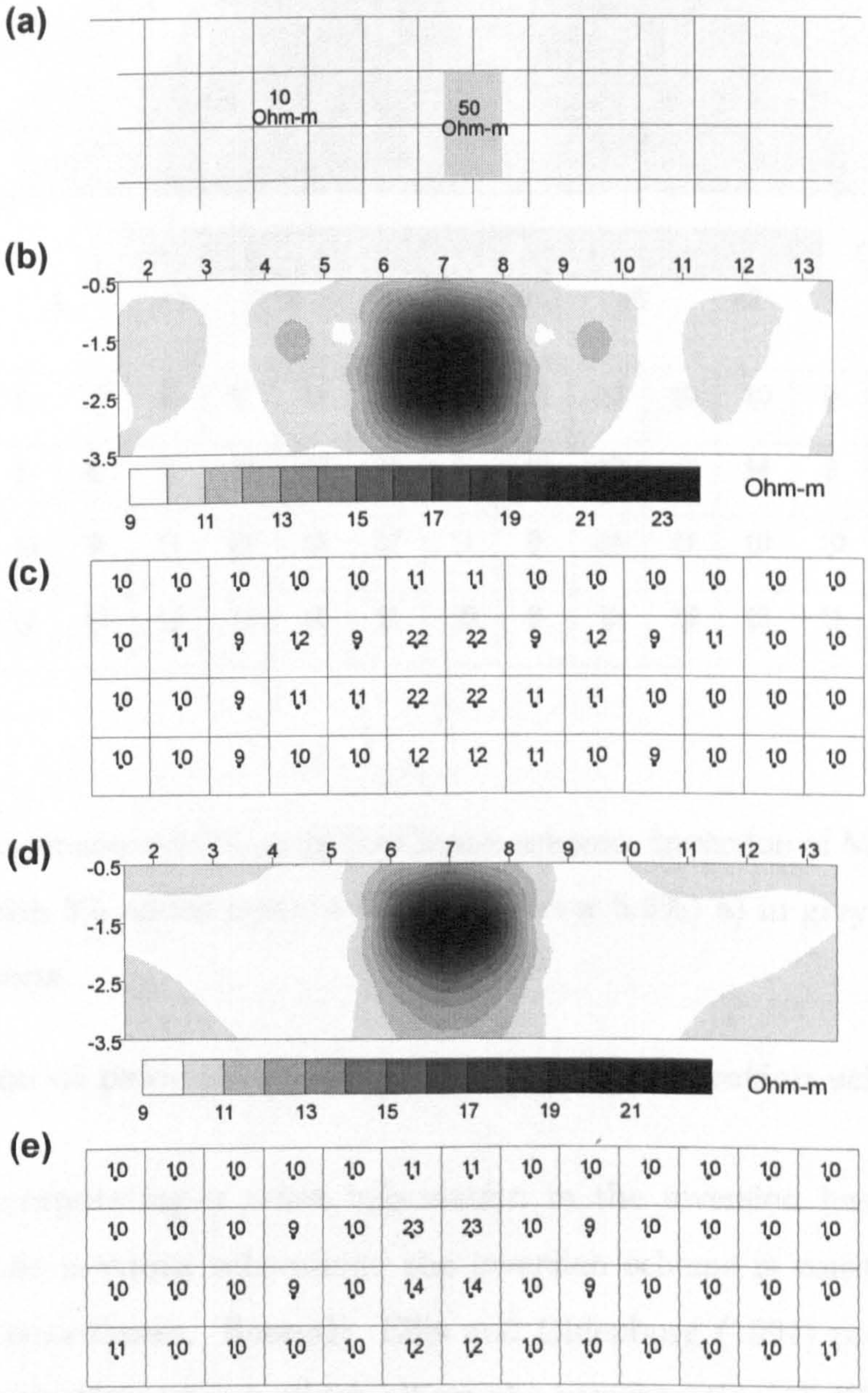
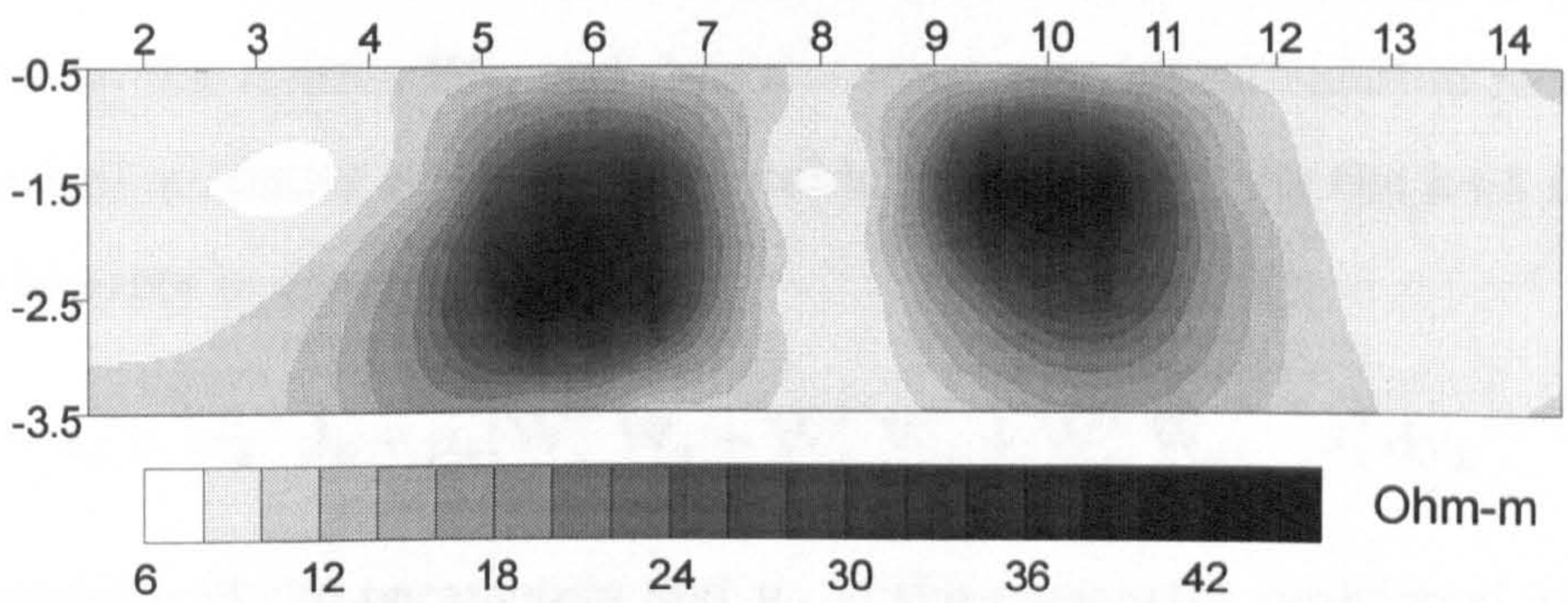


Figure 6.12: a) The model used to produce synthetic the data. Inversion results (dipole-dipole data) using the Occam method: b) in a grey scale form, c) in a post map form. Inversion results (dipole-dipole data) using Marquadt's method: d) in a grey scale form, e) in a post map form.

(a)



(b)

9	10	10	11	11	10	11	10	10	10	10	9	9	11
9	8	7	12	32	35	9	9	42	42	14	9	8	9
10	9	11	21	45	37	11	9	24	31	16	10	8	9
10	11	13	18	16	11	10	8	13	16	13	11	10	10

Figure 6.13: Inversion results using the Occam scheme. Inversion of Model 1 (dipole-dipole data with 5% added noise, 4 iterations, error 5.6%) a) in grey scale form, b) in post map form.

Incorporation of prior information using a regularization scheme

A way for incorporating *a priori* information in the inversion has been demonstrated into the previous subsection: the inversion scheme is constrained with *a priori* model smoothness. Recently, Ellis and Oldenburg (1994) presented a constrained minimization scheme which allows the incorporation of different types of prior information. In their scheme the objective function $\Phi(\mathbf{x})$ to be minimized has the form $\Phi = ||\mathbf{W}_m\,\mathbf{dx}||^2$ where $\mathbf{W}_m$ is a weighting matrix and $\mathbf{dx}$ is the resistivity correction vector. The objective function is defined as a sum of three terms.

$$\begin{aligned}\Phi &= ||\mathbf{W}_m\mathbf{dx}||^2 = \phi_s + \phi_x + \phi_z \\ &= ||\mathbf{W}_s\mathbf{dx}||^2 + ||\mathbf{W}_x\mathbf{dx}||^2 + ||\mathbf{W}_z\mathbf{dx}||^2 = \mathbf{dx} \\ &= \mathbf{dx}^T(\mathbf{W}_s^T\,\mathbf{W}_s + \mathbf{W}_x^T\,\mathbf{W}_x + \mathbf{W}_z^T\,\mathbf{W}_z)\mathbf{dx} \qquad (6.44)\end{aligned}$$

This is equivalent to the smoothness constraint (see section 6.1.5): the matrix $\mathbf{C}$ is here replaced by the matrix $\mathbf{W}_m$ and thus the constrained minimization results in a algorithm which is similar to that of Constable et al. (1989). At the k+1 iteration the model estimate is given by

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \left[\mathbf{J_k}^T \mathbf{J_k} + \mu_k(\mathbf{W}_s^T \mathbf{W}_s + \mathbf{W}_x^T \mathbf{W}_x + \mathbf{W}_z^T \mathbf{W}_z\right]^{-1} \mathbf{J}_k^T \mathbf{dy}_k \qquad (6.45)$$

If n is the number of the parameters and p_{ij} is the parameter positioned at the i_{th} layer and j_{th} column of a 2-D parameter grid then

- $\mathbf{W}_s$ is a diagonal n x n matrix which controls the model's amplitude deviation from the current estimate and has elements $f_{ij}^s \sqrt{\Delta x_{ij}\, \Delta z_{ij}}$ where Δx_{ij}, Δz_{ij} is the length and thickness of the ij_{th} parameter. Note that f_{ij}^x can be equal to $1/\sigma_{ij}^p$ where σ_{ij}^p is the *a priori* variance that the resistivity of the ij parameter can have with respect to the initial model.

- $\mathbf{W}_x$ is a n x n matrix which controls the model's variation in the x direction (horizontal roughness) with elements $\pm f_{ij}^x \sqrt{\Delta z_{ij}/rx_i}$ where rx_i, is the distance between the centres of adjacent parameters in the i_{th} parameter layer.

- $\mathbf{W}_z$ is a n x n matrix which controls the model's variation at the z direction (vertical roughness) with elements $\pm f_{ij}^z \sqrt{\Delta x_{ij}/rz_j}$ where rz_j, is the distance between the centres of adjacent parameters in the j_{th} parameter column.

Note that if all the factors f_{ij}^s, f_{ij}^x, f_{ij}^z are set to be equal to 1 for every parameter ij then the scheme becomes the usual smoothness constrained inversion. The flexibility of the scheme lies in the fact that those factors can become spatially variant, allowing the incorporation of any reliable prior information regarding the possible resistivity variations. If, for example, there are indications that the target lies within a particular region, then the factors corresponding to the parameters of this region can be reduced (reduce the penalty for roughness by reducing f_{ij}^x, f_{ij}^z, and simultaneously assign larger uncertainty limits by decreasing f_{ij}^s).

Such an example was presented by Ellis and Oldenburg (1994): using reliable prior information they produced a spatial weighting function for reducing the resistivity smoothness in the central reconstruction region of their synthetic topographic data set. Their results are considerably better than that of the smoothness constrained inversion for the same region. On the other hand, they did not present any other synthetic or real data examples.

Incorporation of prior information using Bayesian statistics

A more "elegant" procedure for incorporating prior information has been proposed by Tarantola and Valette (1982), Jackson and Matsu'ura (1985). The approach is a stochastic one (as opposed to the deterministic approach demonstrated previously) and is based on Bayesian statistics. The main framework is Bayes's rule which within the context of our problem becomes:

$$P(\mathbf{x}/\mathbf{y}) = \frac{P(\mathbf{y}/\mathbf{x})\ P(\mathbf{x})}{P(\mathbf{y})} \tag{6.46}$$

where: P denotes the probability density function (PDF) and "/" means "conditional". $P(\mathbf{x}/\mathbf{y})$ is the PDF of having a property distribution $\mathbf{x}$ when the observed data are $\mathbf{y}$ (*a posteriori* PDF), $P(\mathbf{y}/\mathbf{x})$ is the PDF of measuring the data $\mathbf{d}$ when the property distribution is $\mathbf{x}$ (represents the solution of the forward problem), $P(\mathbf{x})$ is the prior knowledge about the property distribution and $P(\mathbf{y})$ is a normalizing factor.

The technique involves the definition of an *a priori* model $\mathbf{x}_o$ and a covariance matrix $\mathbf{R}_o$ such that

$$\mathbf{R_o} = \begin{bmatrix} \mathbf{R}_m & 0 \\ 0 & \mathbf{R}_d \end{bmatrix} \tag{6.47}$$

where $\mathbf{R}_d$ is a diagonal covariance matrix which includes the data uncertainties (it is identical to the data error weighting matrix of equation 6.11) and $\mathbf{R}_m$ is a model covariance matrix and is used to weight the model parameters.

The problem can now be formulated within the Bayesian framework: maximize the *a posteriori* probability function (see equation 6.46): in other words, find the model which is both consistent with the measured data and the prior information. Tarantola and Valette (1982) showed that this is equivalent to minimizing the following function:

$$S = (\mathbf{d} - \mathbf{f}(\mathbf{x}))^T \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{f}(\mathbf{x})) + (\mathbf{x} - \mathbf{x_0})^T \mathbf{C}_m^{-1} (\mathbf{x} - \mathbf{x_0}) \qquad (6.48)$$

For the non-linear case the minimization yields an iterative algorithm. For the resistivity case the model at the k+1_{th} iteration is given by (Park and Van, 1991)

$$\mathbf{x}^{k+1} = \mathbf{x}^k + [\mathbf{J}^T \mathbf{R}_d^{-1} \mathbf{J} + \mathbf{R}_m^{-1}]^{-1} [\mathbf{J}^T \mathbf{R}_d^{-1} (\mathbf{dy})^k + \mathbf{R}_m^{-1} (\mathbf{x}_0 - \mathbf{x}^k)] \qquad (6.49)$$

Note that in the right hand side of equation 6.49 a term exists which explicitly involves the *a priori* model $\mathbf{x}_o$. Another key factor for incorporating prior information is the covariance matrix $\mathbf{R}_m^{-1}$. It can be a diagonal matrix, with elements representing the desired variance that each parameter should have relative to the starting model. A very small variance for one parameter, for example, results in keeping it practically constant. In this way, different prior weights can make the inversion more sensitive to some parameters. Further, an *a priori* smoothness can be introduced by using the off-diagonal terms of $\mathbf{R}_m^{-1}$ which couple adjacent parameters (Park and Van, 1991). This can be done in a fashion similar to the construction of the smoothness matrix in Occam inversion.

Pous et al. (1987) presented successful applications of this technique for the 1-D resistivity case. The technique was also used by Park and Van (1991) and Zhang et al. (1995) for inverting 3-D resistivity data. Park and Van assigned equal weighting to all parameters (without including smoothness) and though they started with a realistic initial model they faced instability. They stabilized the procedure by adding a constant to the diagonal elements (similar to the Marquadt damping factor -see section 6.1.4). Zhang et al.(1995) used an identical algorithm. This is indicative that the initial model on its own cannot guarantee stability.

Is Inversion with prior information practical ?

Although there is no doubt about the validity of the schemes which use prior information several questions about its practical applicability arise.

The application of general prior information (such as smoothness) which derives from the physics of the problem is broadly justified. On the other hand other types of prior information and, in particular, the incorporation of spatially varying uncertainty limits on the model parameters is quite arguable. In many cases this type of information does not exist but, most importantly, even if it is available, it is not reliable: the possibility/suspicion that a resistivity anomaly is more likely to exist in a particular region does not give any guarantee that there are not significant resistivity perturbations - which have left their "signature" in the data - in the remaining areas.

Tatantola and Valette stated that more often the *a priori* information will simply be obtained by putting "reasonable" error bars around a "reasonable" central value. Even if a "reasonable" starting resistivity can be established how can a "reasonable" error bar be defined knowing that resistivities can vary over several orders of magnitude? In the majority of real cases prior information is not reliable and therefore many (how many ?) possible error bars might have to be tested before a final interpretation is produced - but this is not practical. To quote Pedersen and Rasmunssen (1989):

> "We would argue that the inclusion of *a priori* information (in the form of elastic uncertainty limits on the model parameters) require one further step of interpretation. Not only do we have to worry about the reliability of data errors, but we also have to speculate as to whether the *a priori* constraints are reasonable."

6.1.7 The Maximum Entropy technique

The maximum entropy method (MEM) is a constrained optimization technique which has been used mainly for handling linear inverse problems, but can be implemented easily for non-linear problems as well (Andrews and Hunt, 1977). It has been widely used in the image restoration field (Gull and Skilling, 1984a,b). Limited use of the technique has been reported in the geophysical field (Robinson, 1985; Christakos, 1991 ; Szymanski and Dittmer, 1992). Preliminary results of its application in the resistivity case have been presented by Szymanski et al. (1994).

The word "entropy" in this case signifies the negative information content of a probability distribution. The basic idea of the MEM optimization is to find the least biased optimum probability distribution (the resistivity for example) based on the existing information. Thus, any solution which fits the data should at the same time maximize the entropy. In that sense the MEM solutions are inherently smooth (Hanson, 1987): no unnecessary complexity (which is not justified by the data and the prior information) is added to the model.

The crux of the MEM is the system independence axiom (Skilling, 1988): if a proportion k of population has a certain property then it is reasonable to assume that any sub-population (which can have this property) has the property in the same proportion k. The consequence of this requirement is that the "best" set of proportions k_j, $(j = 1, 2, ...n)$ for n parameters must be obtained by maximizing the entropy. The entropy of a distribution $\mathbf{g}$ is expressed in a discretized form as

$$E(\mathbf{g}) = \sum_{i=1}^{n} \left[g_j - g_j^o - g_j \log \left(\frac{g_j}{g_j^o} \right) \right] \tag{6.50}$$

where g_j is the unknown distribution , g_j^o is the *a priori* estimate of g_j and n is the number of the parameters. For the resistivity case $E(\mathbf{g}) = E(\mathbf{x})$, $g_j = x_j$ and $g_j^o = x_j^o$ Note that in equation 6.50 a term which explicitly involves the initial *a priori* model $\mathbf{x}_o$ is included.

The optimization problem is now defined as (Gull and Skilling, 1984b): find the

model which has the maximum entropy and at the same time fits the data. In other words, minimize $E(\mathbf{x})$ under the constraint that the sum of square errors $S(\mathbf{x})$, (see equation 6.2, or the weighted sum of square errors $S(\mathbf{x})_w$ - in the case where data error information is available) is minimized. This type of minimization is known as "historic" maximum entropy [13]. The minimization can be achieved by means of a Lagranian multiplier. In the non-linear case the minimization involves the iterative solution of a series of linear problems.

Inversion examples

The MEM was applied to the 2-D inversion of earth resistivity data. The examples presented here are from the work of Szymanski et al. (1994) who combined the MEMSYS 3 package[14] with FEM forward modelling and the adjoint equation technique for calculating the Jacobian matrix.

As applied here, the algorithm involves the solution of a series of linear problems. For every linear problem the optimum resistivity correction vector is obtained by means of a conjugate gradient matrix inversion algorithm and of a search procedure for obtaining the optimum Lagranian multiplier (the philosophy is identical to the 1-D line search involved in Occam inversion - see section 6.1.5).

Here, preliminary results are presented from the reconstruction of the dipole-dipole data set for Model 1 (see Figure 6.1). Inversions were carried out for noise-free data (Figure 6.14a) and data corrupted with 5% additive random noise (Figure 6.14b). In both cases a homogeneous ground was the *a priori* model. The results of Figure 6.14a indicate that the main features are successfully restored, with the deeper areas somewhat less well defined. In the inversion of the noise-corrupted data set the main

[13] A more powerful maximum entropy formulation based on Bayesian statistics -known as "classical" maximum entropy- has been suggested (see Skilling and Gull (1989) for example). Since this technique was not tested here it will not be discussed.

[14] This is a collection of Fortan 77 routines for MEM image reconstruction developed by Maximum Entropy Data Consultans Ltd.

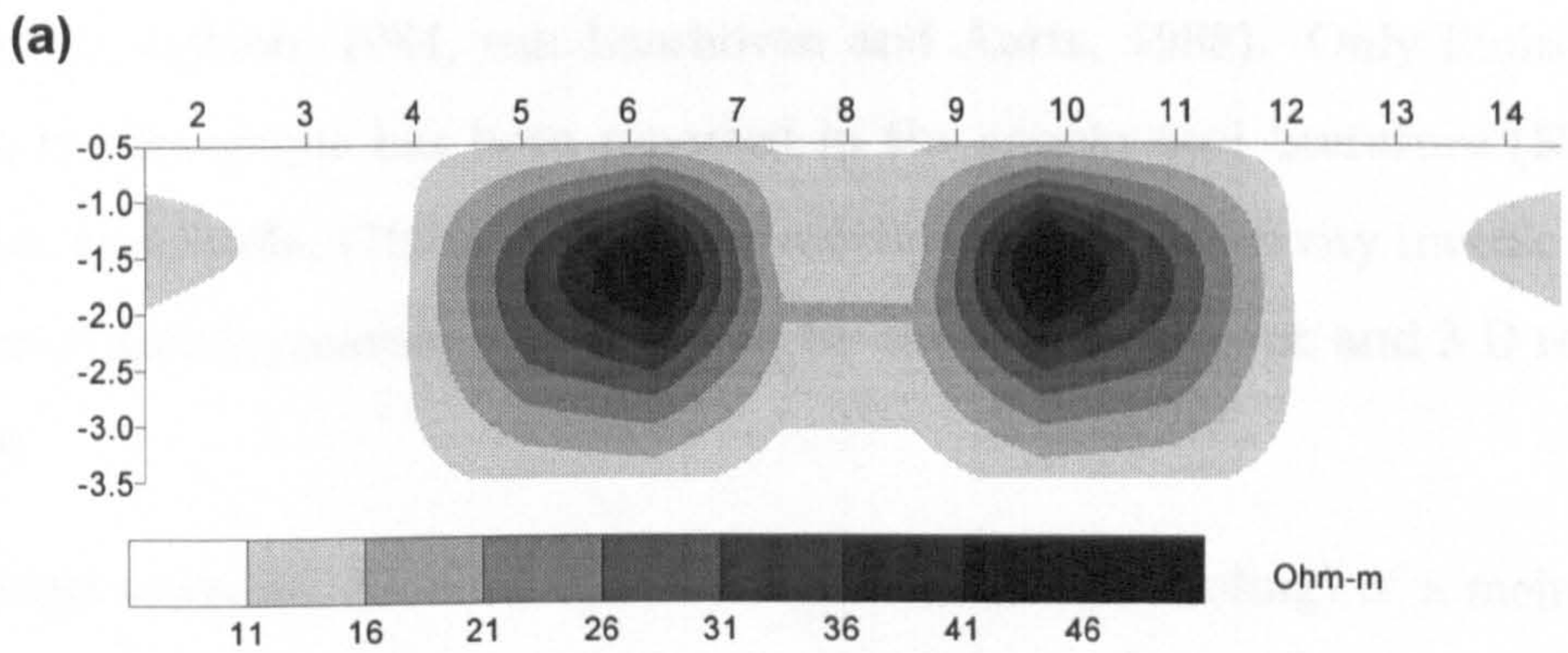


Figure 6.14: Maximum entropy reconstructions of the dipole-dipole data set for Model 1: a) noise-free data, b) data corrupted with 5% noise.

features remain identifiable and the entopic regularization term is very effective at avoiding spurious noise artifacts in the resistivity section.

The results -although preliminary - indicate that MEM can be successfully applied into the 2-D inversion of earth resistivity data. Further understanding of the MEM mechanisms and tests with more complicated synthetic data, but most importantly, with real data are essential for establishing the applicability of the technique.

6.1.8 The Simulated Annealing technique

Simulated annealing (SA) is a stochastic non-linear optimization technique which belongs to the class of methods known as Monte-Carlo methods. It is based on ideas from statistical mechanics, thermodynamics and multivariable probability theory

(Bonomi and Lutton, 1984, van Laarhoven and Aarts, 1988). Only limited application of the technique has been reported in the geophysical literature (Rothman, 1985). Sen and Stoffa, (1993) used SA for solving the 1-D resistivity inverse problem and Dittmer and Szymanski (1995) used it for inverting magnetic and 2-D resistivity data sets.

As the name suggests, SA simulates the solidification (annealing) of a melted solid: the melted solid in its initial state has a large amount of energy and its molecules are positioned randomly, but as it cools the energy is gradually lost and its molecules become less mobile. If the cooling is "gradual enough" its final molecular structure is the one that holds the less possible energy.

The analogy to the inverse resistivity problem can be easily established: the solid becomes the parametrized space and its molecules the parameters. The system's energy can be defined as the sum of square errors between the observed and calculated data S($\mathbf{x}$) (see equation 6.2). Gradual minimization of the energy ("cooling" of the system) will produce the solution vector. The "cooling" procedure can be simulated by the following iterative procedure (Dittmer, 1994):

- Define an energy (cost) function, for example, $S(\mathbf{x}) = \mathbf{dy}^T\,\mathbf{dy}$. Note that terms involving smoothness or other *a priori* constraints can be added to the cost function by the use of Lagranian multipliers.
- Select randomly a parameter i and have its resistivity perturbed by a random (positive or negative) amount Δx_i such that $0 < x_i \pm \Delta x_i \leq \Delta x_{\max}$ (where $\Delta x_{\max}$ is the maximum allowable perturbation).
- Calculate the new forward response $f(\mathbf{x} + \Delta x_i)$ and the new energy.
- If the energy function is reduced (the misfit error is reduced) the resistivity change is always accepted . To determine whether resistivity changes that cause an increase of energy are accepted a control temperature, T, is generated and an algorithm

known as the Metropolis algorithm is consulted:

$$\text{Accept change} \begin{cases} \text{Yes} & \text{if } \exp(-\frac{\Delta S}{T}) \geq R \\ \text{No} & \text{if } \exp(-\frac{\Delta S}{T}) < R \end{cases} \tag{6.51}$$

Where ΔS is the change in energy and R is a random number between 0 and 1.

- While the algorithm is progressing the control temperature T (as well as the maximum allowable perturbation $\Delta x_{\max}$) is gradually reduced to "freeze" the system (the reduction takes place after a set amount of iterations by an amount decided by the user (i.e. 90% reduction after every 500 iterations).
- The procedure is repeated until practically no changes in energy are accepted or the misfit is just above the observation error.

One of the most important components of the SA algorithm is the Metropolis algorithm: the fact that resistivity changes which increase the misfit can be accepted is essential for preventing the technique from being trapped in local minima. In this way the global searching character of the method is guaranteed, providing that the temperature profile is not to steep.

Other advantages of the technique is that its progress is not determined by the conditioning of the controlling matrices, and that it enables the search of a greater proportion of the possible model parameter configurations (Dittmer and Szymanski, 1995). On the other hand, it has to be noted that it is significantly slower (almost by a factor of 30) than the techniques involving matrix inversion (Tsourlos et al. 1995)[15]. Furthermore, its use involves many quantities (such as the control temperature, the maximum perturbation, etc.) that the user has to define arbitrarily, and this is a practical disadvantage.

[15]Dittmer(1994) also proposed an accelerated version of the SA, but no tests with this algorithm were made by the author.

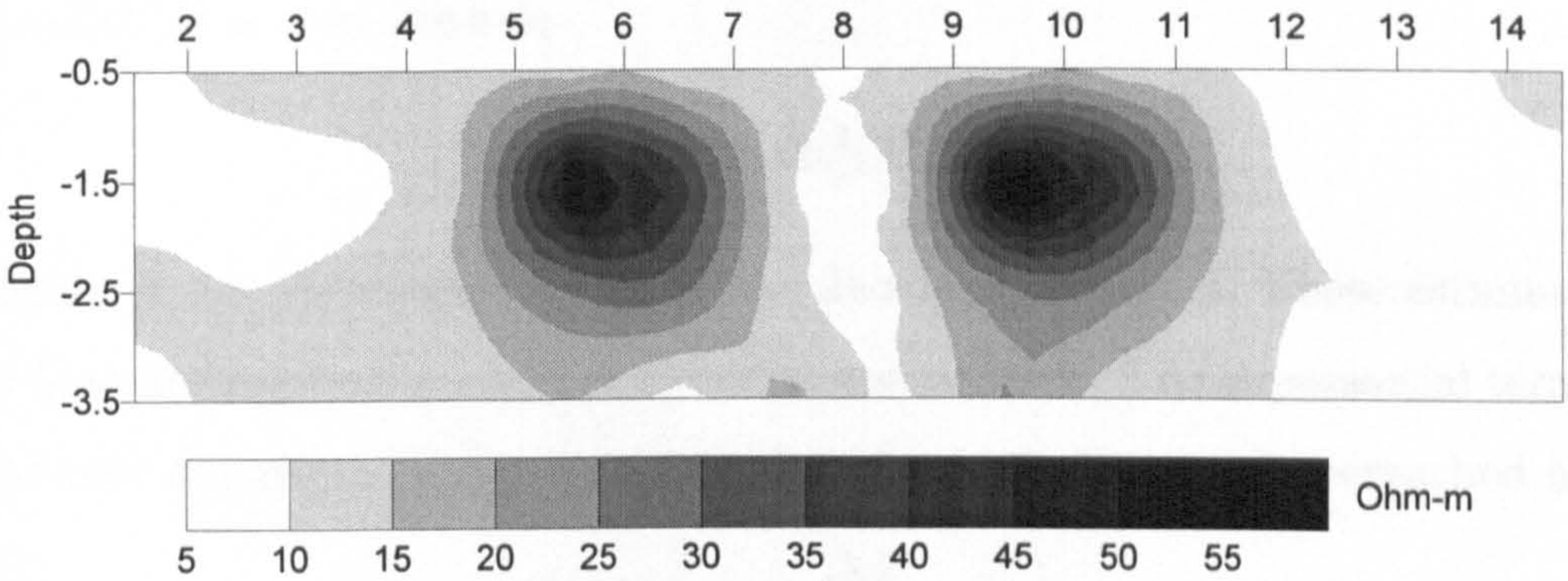


Figure 6.15: Simulated annealing reconstruction of the dipole-dipole data set for Model 1.

Inversion examples

The SA optimization was used for the 2-D inversion of resistivity data. The forward modelling and parametrization schemes are identical to those used for testing the previous techniques. The core SA routine was written by Dr. Jonathan Dittmer.

The initial tests with dipole-dipole data indicated that the Metropolis algorithm was favouring changes at the top layers more than at the bottom. This coincides with what Dittmer (1994) observed for the magnetic SA inversion and a similar justification applies: the parameters at the top layers will have a higher contribution in the "cost" function than those at the bottom layers (in line with the ill-posed nature of the problem). Thus, changes in upper layers are more likely to be accepted by the Metropolis algorithm (see equation 6.51).

To tackle this problem, the author introduced the following modification: The Jacobian matrix was calculated for a homogenous ground. A quantitative estimate of the energy difference δs_j that a "small" resistivity perturbation δx_j of a parameter j will produce can be obtained by

$$\delta s_j = (\mathbf{J}\mathbf{dx}_j)^T (\mathbf{J}\mathbf{dx}_j) \tag{6.52}$$

where $\mathbf{dx}_j$ is a vector with all elements but its j_{th} ($dx_j = \delta x_j$) set to zero. Thus

equation 6.52 is equivalent to

$$\delta s_j = \delta x_j^2 (\mathbf{J}_j^T \mathbf{J}_j) \tag{6.53}$$

where $\mathbf{J}_j$ is the j$_t h$ column vector of the Jacobian matrix $\mathbf{J}$. These estimates were used as weighting factors on the control temperature T. The exponential term in the Metropolis algorithm (equation 6.51) when the j$_{th}$ parameter is perturbed becomes

$$\exp\left(\frac{\Delta S}{T\,(\delta s_j / \delta s_{\max})}\right) \tag{6.54}$$

This weighting scheme produced more balanced changes.

A further modification was introduced to balance the fact that the random scheme for changing the parameter resistivity does not favour conductive bodies: if the resistivity of a parameter is small (i.e. 5 Ohm-m) and the maximum allowable perturbation $\Delta x_{\max}$ is relatively large (i.e. 20 Ohm-m) (note that this is quite a realistic situation) then the possibility that the next resistivity perturbation change will be positive is much higher. To avoid this type of problem the usual logarithmic weighting was introduced (see section 6.1.2).

In Figure 6.15 the SA inversion for Model 1 after 6000 iterations can be seen[16]. The reconstruction delineates the shape of the targets successfully. However, (even though the data set was noise free) some reconstruction artifacts related to the bottom side parameters appeared. They are due to the mechanism of the SA: as the system "cools" and the control temperature drops, the resistivity of the parameters of the top layers settles. In the later stages of the cooling it is effectively only the parameters with a small contribution to the misfit error (energy function) that are allowed to change (despite the fact that theoretically all parameters can change). These "unimportant" parameters are related to the bottom and side edges of the studied area. Thus, even if the total energy is low it would take several attempts until these parameters were finally corrected. In this sense the inversion presented in

[16]Note that for the SA one iteration involves effectively only the solution of the forward problem. On the other hand, one iteration signifies a successful resistivity change - the unsuccessful attempts are not included

this figure is somehow "premature", and cannot improve while the upper parameters are fixed.

6.1.9 Discussion and Conclusions

In this section several non-linear techniques used in the solution of the 2-D resistivity inverse problem were presented. It was shown that ill-conditioning combined with data-errors generates unstable solutions. As far as the tested techniques are concerned the following can be said:

- Marquadt's method produced very good results when noise-free data were tested, but in the case of noisy data spurious noise-related artifacts appeared. More generally, the problem with Marquadt's method is that it is based on a mathematical "trick" which produces stability but cannot guarantee physically realistic solutions. Apart from defining the initial model, the user cannot have any other control on the procedure, and thus the results will be highly dependent on the (successful or not) choice of the initial model.

- Smoothness constrained inversion, on the other hand, produces a solution which has the general properties that the user has *a priori* selected. Requiring the solution to be smooth is physically realistic and particularly suited to the case of 2-D inversion (as was demonstrated by examples). The tests with noisy data indicated that the smoothness constraint is successful in avoiding noise related artifacts. Furthermore, the technique proved particularly robust and although it did not produce the "exact" solutions (it would be unrealistic to expect perfect results from any inversion scheme where inaccurate observations or an ill-conditioned system are involved) the results were quite close to reality.

Elis and Oldenburg (1994) suggested a case where smoothness constrained inversion did produce totally smooth reconstructions which had no resemblance to the reality but their example is a special case with no general value. We noticed a similar

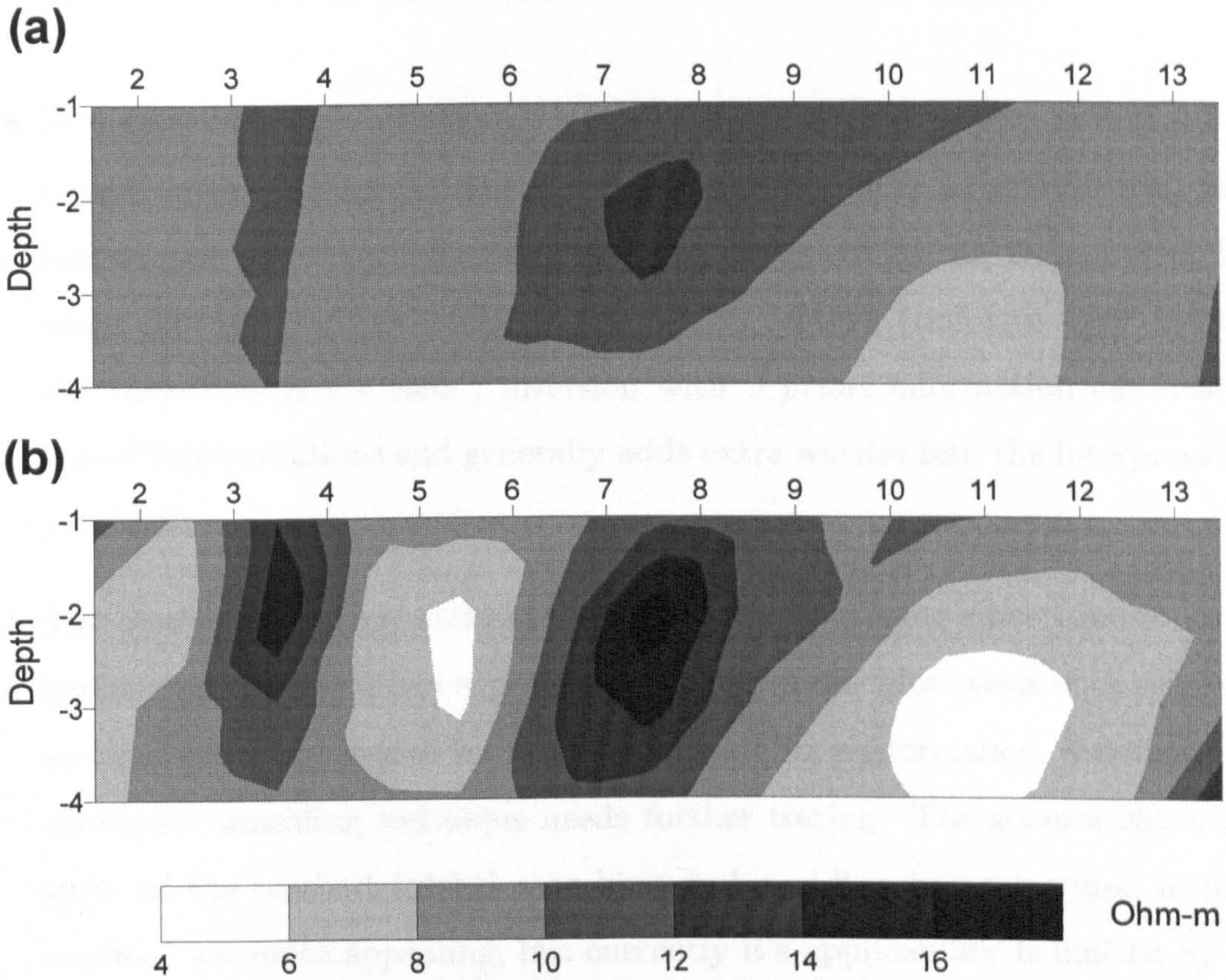


Figure 6.16: Reconstruction of noisy data (40 % noise). a) using the Occam method, b) using the Marquadt's method.

behaviour for some of the tested cases: with extremely noisy data the technique tended to smooth the solution in such a way that all features are suppressed and the final result did not have much resemblance to the reality, however under no circumstances were features produced that were not there. Such an example is shown in Figure 6.16. 40% noise was added to the data set which corresponds to the model of Figure 6.12a. The Occam inversion (Figure 6.16a) delineates the target slightly, while for Marquadt's method the prism is successfully recovered (Figure 6.16b) but the reconstruction it was full of major artifacts. Even though both inversion techniques were unsuccessful from a detailed interpretational point of view, the Occam inversion is preferable since it is quite reassuring to know that the inversion procedure is unlikely to produce spurious features. In other words,

"no-information" is better than "bad-information".

- Inversion with *a priori* information is theoretically even more powerful than Occam inversion: apart from general constraints such as smoothness, prior information regarding the spatial variation of resistivity can be included. But when this type of prior information is not reliable (unfortunately most of the times this is the case) inversion with *a priori* information can lead to biased interpretations and generally adds extra worries into the interpretation procedure.

- The maximum entropy technique looked quite promising since (due to its inherent smoothness) it can cope well with noisy data. More tests with synthetic and real data are needed in order to evaluate its performance. Similarly, the simulated annealing technique needs further testing. The general characteristics of the method (global searching and avoiding being trapped in local minima) are quite appealing, but currently it's applicability is limited by the enormous computational loads that it involves.

In conclusion, it is believed that currently the smoothness constrained inversion is the most preferable technique for practical data interpretation. Although far from perfect it comprises numerous advantages: **a)** it is physically reasonable, **b)** it is robust, **c)** it can cope well with noise and does not produce artifacts, **d)** it produces solutions which have properties that the interpreter has chosen and are not a product of an arbitrary initial choice. For these reasons the smoothness constrained inversion was chosen as the platform for developing a fast 2-D inversion algorithm. This algorithm, together with further examples (with synthetic and real data) of the smoothness constrained inversion, will be presented in the following section.

6.2 A fast smoothness constrained algorithm for the 2-D inversion of earth resistivity data

In this section a fast smoothness constrained algorithm for the 2-D inversion of earth resistivity data is presented. The procedure is fully automated and accelerated by the use of a Quasi-Newton update of the Jacobian matrix. In the following subsections the need for a fast algorithm is discussed and a detailed presentation of its features is given. Further, its performance is compared to other techniques on the basis of synthetic and real examples. Finally, extensive examples of its application to real data are presented and discussed.

6.2.1 The need for a fast non-linear inversion algorithm

The need for increased speed in the inversion procedure derives from the advent of automatically controlled resistivity meters. Most of the automated systems are able to control a large number of electrodes (i.e. 20-100) and in this way not only the amount of collected measurements, but also the speed of the data collection is increased. Eventually the geophysicists ends up with a large number of high resolution data sets.

A rapid initial interpretation of the data sets is valuable and will help to check if the right survey settings are chosen, and can aid in the design of further surveys in the area of interest. Traditionally, the "approximate" inversion techniques are used for this purpose whilst "accurate" techniques are usually used for final processing/interpretation after the survey has ended.

Therefore, it is clear that there is a scope for the development of a fast fully non-linear algorithm which could cope with the increased amount of data and could be used as a tool for more accurate initial data interpretation. Furthermore, the "accurate" algorithm can still be used during the final interpretation stages.

For most of the iterative schemes presented in this chapter the total time for one iteration is effectively divided between three basic procedures: the calculation of the Jacobian matrix, the calculation of the forward model, and the matrix inversion. For a typical data set obtained using an automated procedure (i.e. 40 electrodes, 300 measurements) the calculation of the Jacobian matrix takes approximately 70% of the iteration time and the remaining time is equally divided between the calculation of the forward model, and the matrix inversion. As the data set gets larger the matrix inversion starts getting more time-consuming and for really large data sets[17] (i.e. >1000 measurements) it becomes dominant.

Since for a typical size of data set the matrix inversion is not that time-consuming, it follows that the inversion procedure can be accelerated significantly by avoiding the direct calculation of the Jacobian at every iteration[18]. This can be achieved by the use of the Quasi-Newton techniques which will be described in the following section.

6.2.2 Quasi-Newton techniques

Quasi-Newton (QN) (or variable metric) techniques are a class of non-linear optimization methods which seek to approximate the Jacobian at each iteration instead of calculating it from scratch. This type of method is like the Gauss-Newton technique (see section 6.1.1) except that the Jacobian matrix **J** is approximated by a matrix **H** which is corrected and updated from iteration to iteration (Fletcher, 1987).

There is a huge literature on QN techniques since, as Fletcher (1987) suggests, the possibilities of QN formulae are endless. It is worthwhile mentioning that QN

[17]This type of data set derives from extensive roll-on surface surveys or, more typically, from crosshole or borehole-to-surface surveys

[18]For very large data sets fast matrix inversion techniques can also be used (see Zhang at al. (1995) for an example).

formulas have been produced for constrained optimization problems (Gill and Muray, 1974). Note that most of the QN techniques are also able to produce an approximate inverse of the Jacobian. Use of the QN techniques in the 2-D resistivity inverse problem has been made by Shima (1990)[19] who used Powell's algorithm (Powell, 1970). This algorithm combines a QN Jacobian update and Marquadt's method. Further, Loke and Barker (1994) gave a short presentation of a QN scheme for inverting earth-resistivity data.

To indicate how the QN techniques work, consider a measurement vector $\mathbf{y}$ and an initial property distribution vector $\mathbf{x}_o$. If $\mathbf{J}_o$ is the Jacobian matrix and $\mathbf{f}(\mathbf{x}_o)$ is the forward modelling response then one iteration of any optimization technique which involves matrix inversion will produce a parameter correction vector $\mathbf{dx}_o$. The forward response for the new estimate will be $\mathbf{f}(\mathbf{x}_o + \mathbf{dx}_o)$. The target is to find an approximate expression $\mathbf{B}_1$ of the Jacobian $\mathbf{J}_1$ for the new iteration without calculating it from scratch. A simple way is to approximate the new Jacobian in a finite difference sense (recall the perturbation technique -section 4.3.3):

$$\begin{aligned} B_1^{(ij)} &\simeq \frac{\mathbf{f}^i(\mathbf{x}_o + \mathbf{dx}_o) - \mathbf{f}^i(\mathbf{x}_o)}{\mathbf{dx}_o^j} \quad \text{or} \\ \mathbf{B}_1\,\mathbf{dx} &\simeq \mathbf{f}(\mathbf{x}_o) - \mathbf{f}(\mathbf{x}_1) \end{aligned} \tag{6.55}$$

Equation 6.55 is known as the Secant method. One of the problems of this technique is that it does not guarantee uniqueness (Burden and Faires, 1993): there is no indication that the approximate Jacobian $\mathbf{B}$ will affect vectors which are orthogonal to $\mathbf{dx}$ in the same way that the real Jacobian $\mathbf{J}$ does. Brodyen (1965) showed that the unique $\mathbf{B}_1$ can be defined as:

$$\mathbf{B}_1 = \mathbf{J}_o - \{\,\mathbf{J}_o\mathbf{dx}_o - [\,\mathbf{f}(\mathbf{x}_o) - \mathbf{f}(\mathbf{x}_1)\,]\,\}\,\frac{\mathbf{dx}_o}{\mathbf{dx}_o^T\,\mathbf{dx}_o} \tag{6.56}$$

Equation 6.56 can be used to update the estimate of the Jacobian in every iteration.

[19]Shima (1990) had to use a QN update for the Jacobian since he used an integral equation technique (alpha-centres) for his forward model calculations. The integral equation techniques cannot be used to calculate the Jacobian matrix directly.

For the k+1 iteration the approximate estimate is given by

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \{\mathbf{B}_k \mathbf{dx}_k - [\mathbf{f}(\mathbf{x}_k) - \mathbf{f}(\mathbf{x}_{k+1})]\} \frac{\mathbf{dx}_k}{\mathbf{dx}_k^T \mathbf{dx}_k} \tag{6.57}$$

where $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{dx}_k$. Note that from equation 6.56 an updated estimate for the inverse of $\mathbf{B}$ can also be obtained.

One of the disadvantages of the QN methods is that they have superlinear convergence, as opposed to the quadratic convergence of the Newton-like methods. In 2-D resistivity inversion QN techniques typically result in an average increase of 1-3 iterations. For most cases, however, this trade-off is more than acceptable considering the decrease in the amount of computation involved. Only in cases when the iteration time is dominated by the matrix inversion (extremely large data sets) does this become a real disadvantage.

The main disadvantage of the technique is that it is subject to errors involved with the finite difference type of approximation of equation 6.56. One problem is that round-off errors propagate: each new approximate Jacobian $\mathbf{B}_{k+1}$ is a function of the previous (also approximate) Jacobian $\mathbf{B}_k$. This is an extra source of error in the inversion (additional to the observation errors) and might cause instability. It will be shown that was not found to be the case: the smoothness constraint in the inversion scheme prevented the round-off errors from causing unstable solutions.

6.2.3 Algorithm description

The Broyden's QN technique was combined with the Occam inversion scheme in order to produce a fast non-linear algorithm for the 2-D inversion of earth resistivity data. The algorithm uses a FEM algorithm for the forward modelling needs. A general description of the algorithm follows.

Initial Steps

Given a measured data set **y**

a. Define the model parameters

b. Produce the matrix **C** which describes the smoothness pattern of the model

c. Define an initial resistivity estimate $\mathbf{x_0}$ and calculate the model response $\mathbf{f}(\mathbf{x_0})$.

d. Calculate the Jacobian matrix $\mathbf{J_0}$ which corresponds to $\mathbf{x_0}$ and set $\mathbf{B_0} = \mathbf{J_0}$.

e. Set the initial value μ_0 of the Lagranian multiplier.

f. Set the inversion stopping criteria

1. At the k_{th} iteration the resistivity correction vector $\mathbf{dx_k}$ is given by

$$\mathbf{dx_k} = (\mathbf{B_k}^T \mathbf{B_k} + \mu_k \mathbf{C}^T \mathbf{C})^{-1} \mathbf{B_k^T} \mathbf{dy_k} \tag{6.58}$$

   where $\mathbf{B}_k$ is the QN Jacobian estimate which corresponds to the $\mathbf{x}_k$ resistivity distribution, and $\mathbf{dy_k} = \mathbf{y} - \mathbf{F}(\mathbf{x_k})$.

2. Set the new resistivity estimate $\mathbf{x}_{k+1}$ to $\mathbf{x_{k+1}} = \mathbf{x_k} + \mathbf{dx_k}$ and calculate the forward response of the new model $\mathbf{f}(\mathbf{x}_{k+1})$.

3. If one of the stopping criteria are met end the procedure, else find the new QN estimate of the Jacobian using equation 6.57 and go to step 1.

Some of the practical considerations of the algorithm are discussed below

The parametrization scheme

To render the algorithm fully automated a scheme for automatic generation of the parameter space was included. A reasonable parametrization scheme should:

a) Generate a number of parameters which is almost equal to - but not more than - the number of measurements (maximum resolution without having an under-determined system).

b) Be able to cope with any known resistivity array

c) Reduce - as far as possible- the number of irrelevant parameters.

d) Be able to cope with new/unusual array types.

After several sensitivity analysis tests (see section 4.4) the following scheme was produced:

- The parameters are adjusted in a pseudosection-like form: the number of parameters in each layer is symmetrically reduced as depth increases. That is because a sensitivity analysis indicated that parameters at the edges tended to become irrelevant as the depth is increased.
- The number of the parameter layers is set equal to the max n-separation of the measured data set.
- The thickness of each layer is set to be 0.5 of the inter-electrode spacing for every array used. Broadly speaking, this is consistent with the depth of investigation (see discussion in section 2.2.3) for the most commonly used arrays and was verified by the sensitivity analysis tests.
- Each parameter column is positioned between two subsequent electrode positions.

The number of parameters produced by this scheme is quite close to (but always lower than) to the number of measurements (assuming that a full data set is obtained). In Figure 6.17a the resultant parameter mesh for the case of 20 electrodes and a n-separation of 5 for the dipole-dipole, pole-pole and pole-dipole arrays is depicted. Special consideration had to be taken of the Wenner array which involves less measurements in the combined sounding profiling mode (see Figure 6.17b.). Note also that the x dimension of the side and the z dimension of the bottom parameters were set to be quite large (> 8 electrode spacings) to simulate infinite boundaries.

Further, the option of changing the number of the parameter layers and/or redefine the thicknesses of the existing parameter layers was introduced into the scheme. This was made for two reasons:

- There might be missing data points or the measuring scheme could be incomplete. If too much data is missing then the automatic parameter generation scheme (which assumes that is dealing with a full data set) will create more parameters than the existing measurements (under-determined system).

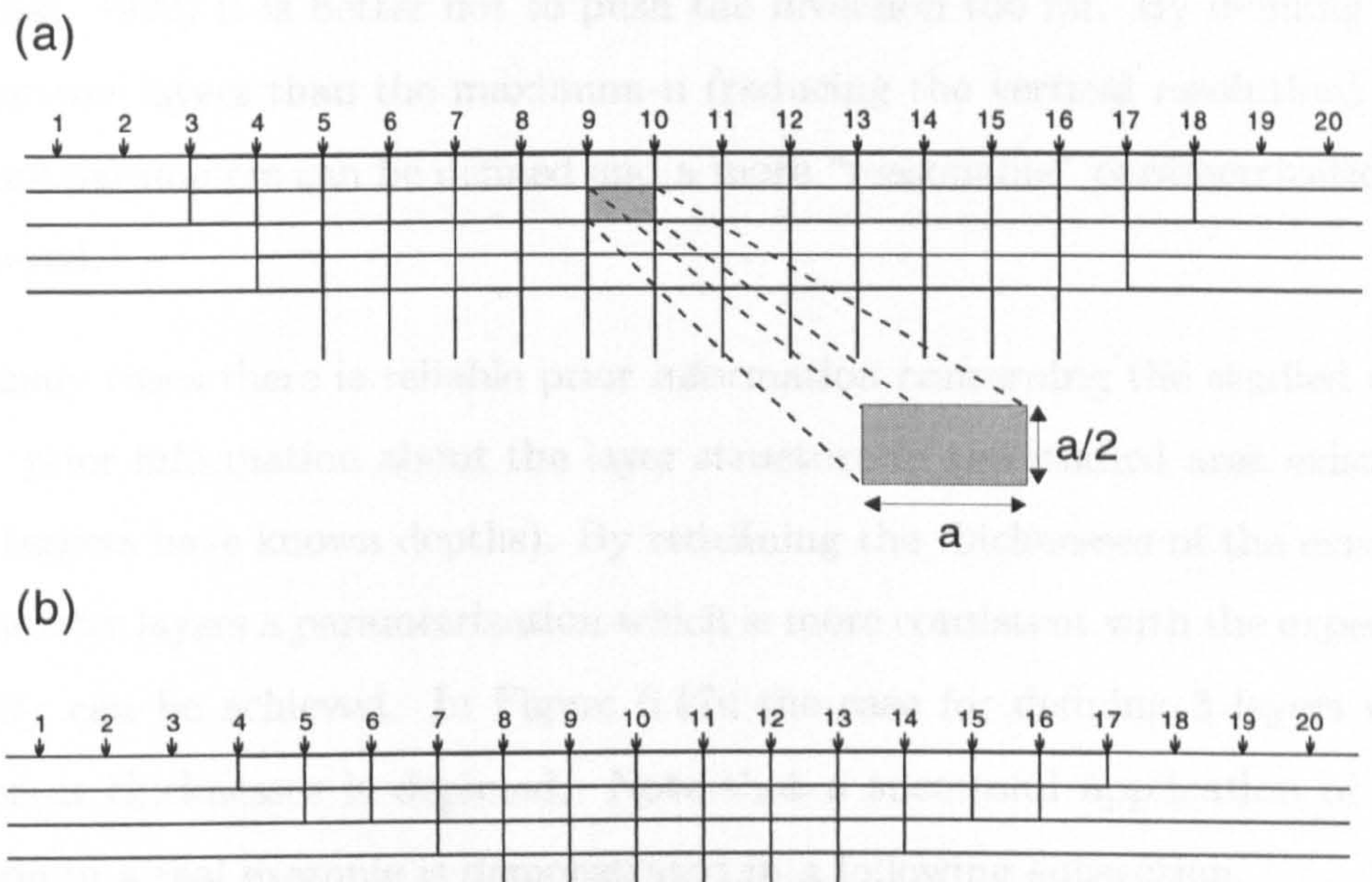


Figure 6.17: The parametrization scheme used by the algorithm: a) the case of dipole-dipole, pole-pole, pole-dipole arrays (20 electrodes, n=5), b) the case of Wenner array (20 electrodes, n=5) c) the case of a redefined number of layers and thicknesses.

Although SC inversion can cope successfully with under-determined systems (Sasaki, 1992) it is better not to push the inversion too far. By defining less parameter layers than the maximum-n (reducing the vertical resolution) less overall parameters can be defined and a more "reasonable" parametrization is achieved.

- In many cases there is reliable prior information concerning the studied area (i.e. prior information about the layer structure in the studied area exists or the targets have known depths). By redefining the thicknesses of the existing parameter layers a parametrization which is more consistent with the expected reality can be achieved. In Figure 6.17c the case for defining 3 layers with different thicknesses is depicted. Note that a successful application of this option in a real example is demonstrated in a following subsection.

Smoothness matrix

As discussed in section 6.1.5 the smoothness matrix **C** describes the smoothness relations between the parameters. The smoothness pattern in this algorithm is a slightly modified version of the scheme proposed by Sasaki (1989). According to his scheme the roughness of the spatial variation of resistivity for a parameter j is given by:

$$\delta\tilde{x}_j = a_j[\delta x_j^E + \delta x_j^W + \delta x_j^N + \delta x_j^S - 4\delta x_j] \tag{6.59}$$

where E, W, N, S indicate the four immediate neighbours of the jth block and a_j represents an empirically defined gradient-amplifying factor. If n is the number of parameters then **C** is a n x n matrix whose coefficients are either a_j, $-4a_j$ or 0.

One possible choice for the factor a is the following: set a_j=1, 1.1 ,1.2 ... for the parameters in the 1st, 2nd 3rd ... layers respectively. In this way the smoothness is increased with depth. This is justified by the fact that resolution decreases with

depth while ill-conditioning increases. Thus the deeper the parameter the more smoothness constrained should be. The detail of this allowance is, however, an arbitrary decision by the operator.

The modified smoothness pattern used in this algorithm is given by

$$\delta\tilde{x}_j = [a_j^x(\delta x_j^E + \delta x_j^W) + a_j^z(\delta x_j^N + \delta x_j^S) - 2(a_j^x + a_j^z)\delta x_j)] \qquad (6.60)$$

where a_j^x, a_j^z are weighting factors which control the roughness in the horizontal and vertical direction respectively. They are adjusted in such a way that they compensate for the possible unequal length δx_j ,and thickness δz_j, of the parameter j. The idea of such a weighting is due to deGroot-Hedlin and Constable (1990) who suggested that if a parameter has unequal dimensions then this is equivalent to having an increased smoothness in the direction of the largest dimension. This can be balanced by assigning an increased smoothness (roughness penalty) to the smaller dimension. In our case the weighting factors become $a_j^x = 1$, $a_j^z = \delta x_j, /\delta z_j$[20]. Finally it has to be mentioned that equation 6.60 gives the ability to assign directional smoothness -i.e. decrease the roughness penalty in the x direction if lateral variations of resistivity are of interest. In Figure 6.18 the matrix **C** for the simple case of 9 parameters can be seen. Note that the x,z dimensions of the parameters are not equal.

The Lagranian Multiplier

In this scheme an empirical way for deciding the Lagranian multiplier at every iteration is used. This scheme was preferred to the 1-D line search procedure (see section 6.1.5) since the latter proved to be quite time-consuming[21] and would change the "fast" character of our scheme. The empirical scheme (which was established

---

[20]Note that for the case of square parameters equation 6.60 becomes identical to 6.59.

[21]Recall that a modest line search needs at least three repetitions of the forward modelling and matrix inversion procedure

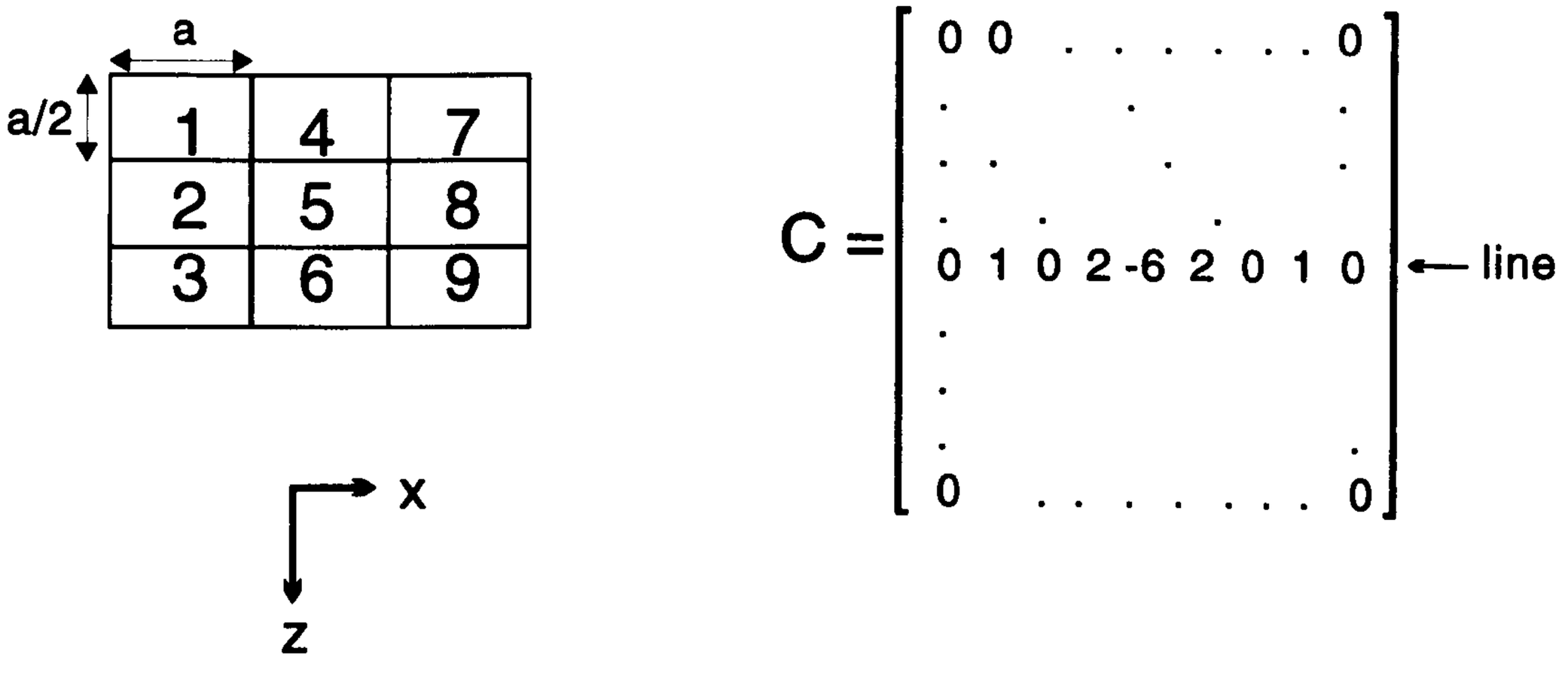


Figure 6.18: The smoothness matrix C (right) for the case of a 9 parameter mesh (left). Note that the x,z dimensions of the parameters are not equal.

after several tests with synthetic and real data) is the following:

$$\mu_k = \begin{cases} \mu_{k-1}/2 & \text{if } k \leq 4 \\ \mu_{k-1} & \text{if } k > 4 \end{cases} \qquad k = 1, 2 \ldots \text{number of iterations} \tag{6.61}$$

This scheme proved quite satisfactory and, in the tested cases, produced inversions very similar to those obtained by the 1-D line search scheme. The only disadvantage of the empirical scheme compared to the 1-D line search is that sometimes it took an extra 1-2 iterations to reach the minimum error. But this is an acceptable trade-off. Actually, as Constable et al. (1987) has suggested, there is no guarantee that the 1-D line search procedure will produce a model that fits the data better. Thus, there is no reason to believe that the empirical scheme is inferior to the 1-D line search scheme.

The only requirement of the empirical scheme is that it needs a user-defined initial Lagranian multiplier value μ_o. A test/indication for choosing a proper value is the following:

If the inverted model changes only slightly during the first 2 iterations then the value of LM is fairly high and it has to be reduced.

If the inverted model changes abruptly during the first 2 iterations, producing un-

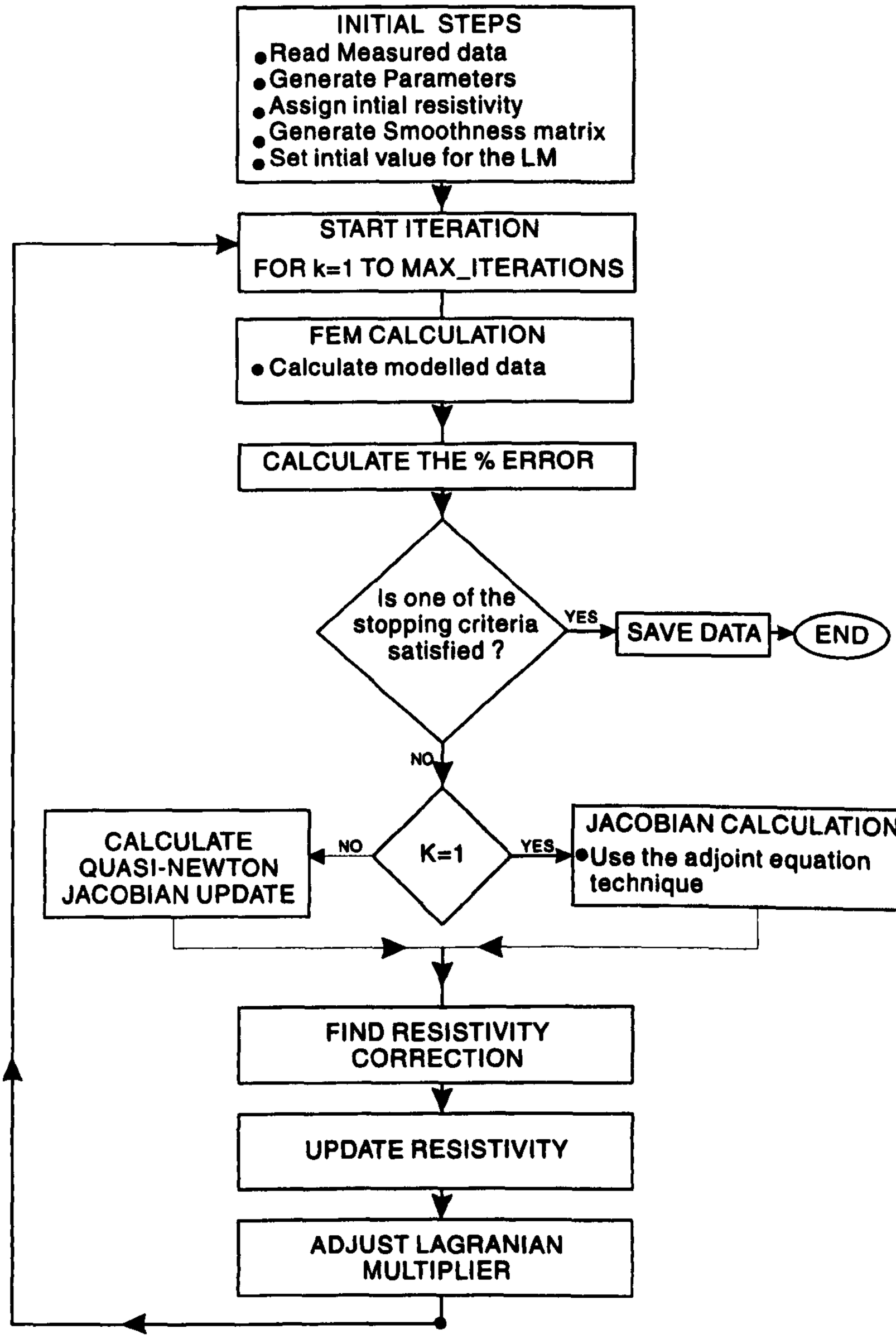


Figure 6.19: A simplified flow-chart of the QN Occam algorithm.

reasonable high or/and low resistivity values, then the value of the LM is fairly low and it needs to be increased.

After testing several models the following empirical values for LM seem to be appropriate for most cases:

- For noise-free synthetic data a value of $\mu_o = 0.05$ was found to be optimum.
- The value $\mu_o = 0.5$ was found to be satisfactory for the vast majority of the tested real data sets. This value has to be increased only if very noisy data sets are inverted.

Stopping criteria

There are four possible stopping criteria that the algorithm could apply depending on whether data noise information is incorporated: After the correction into the resistivity vector is made the % relative RMS error is calculated:

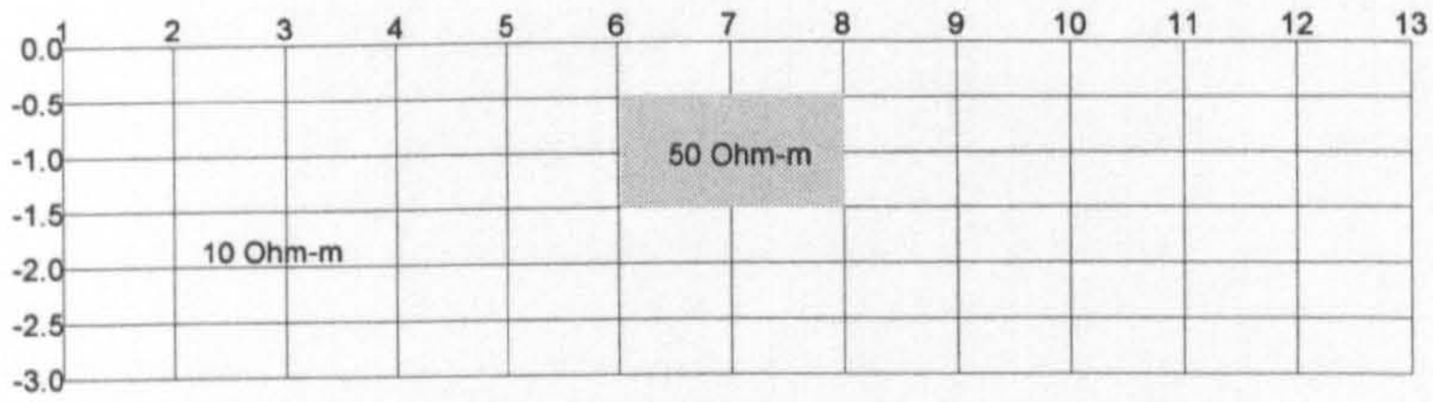
$$\% \text{ RMS error } = 100 \cdot \sqrt{\left[\frac{1}{M}\sum_{i=1}^{M}\frac{(d_{obs}i - d_{calc}i)^2}{d_{obs}i^2}\right]}$$

Where M is the number of measurements, $d_{obs}i$ is the i observed measurement $d_{calc}i$ is the i th calculated measurement. The inversion will stop in the following cases:

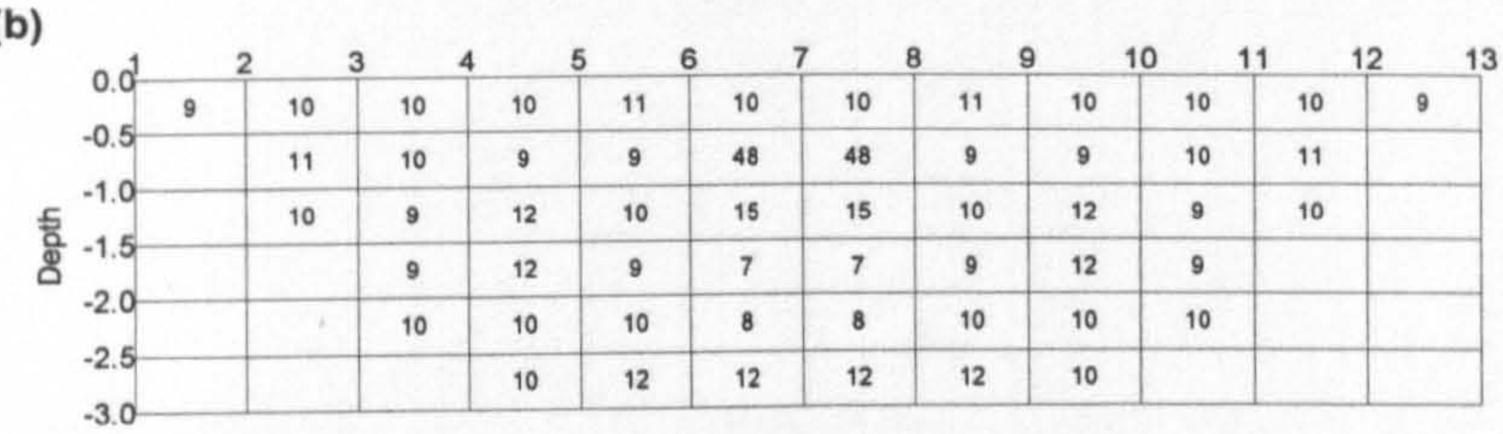
1. **Divergence** The algorithm stops if the error between the real and model data increases (occurrence of divergence). Divergence may occur in rare cases such as extremely noisy data or unsuccessful (very low) choice of LM.

2. **Slow convergence rate** the algorithm stops if the error between the real and model data decreases in a slow rate (less than 5%). Actually, in such a case the inversion procedure could continue but there is the danger that the data will start to fit to noise. In fact, this can also happen even though the convergence rate is not slow.

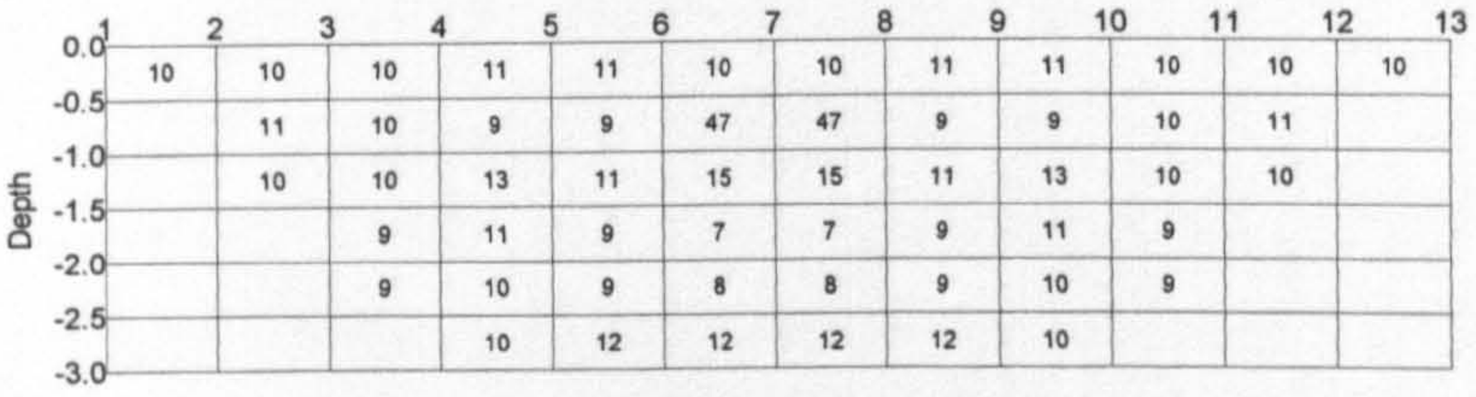
(a)



(b)



(c)



(d)

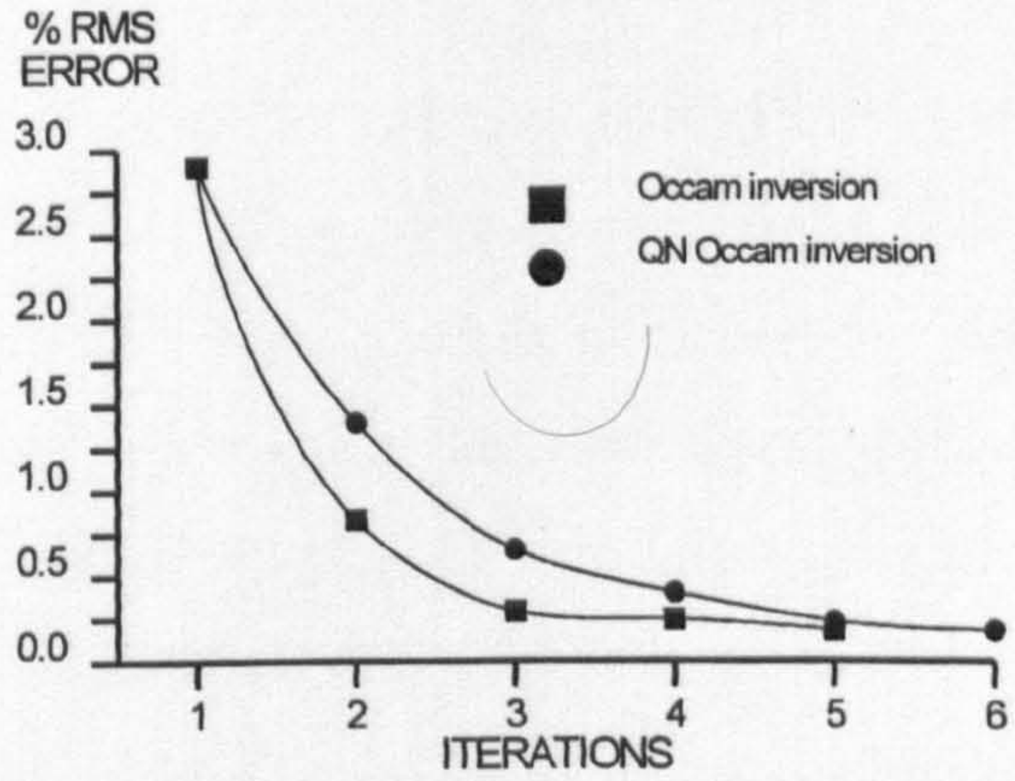


Figure 6.20: Inversion of synthetic dipole-dipole data (noise-free, 15 electrodes, n=5): a) the model used to produce the data. b) Inversion results using the QN Occam method. c) Inversion results using the Occam method. d) The convergence of the two techniques.

3. **The misfit error is below the observation error** This case is activated when data error information is given. If the inversion error is lower than the standard deviation of the data errors then the inversion stops (since data have started to fit to noise) and the resistivity estimate of the previous iterations is the valid result. Note that in such a case the inversion misfit error is $S(\mathbf{x})_w$ (see equation 6.10).

4. **Number of iterations completed** The algorithm could stop if a predefined number of iterations is completed.

6.2.4 Application and evaluation of the algorithm

The described algorithm was applied to a series of synthetic data. The finite element method was used as the forward modelling technique. The initial Jacobian matrix was calculated by the adjoint equation technique. The matrix inversion was performed by the use of the singular value decomposition routine of Press et al. (1987).

Because of the option of assigning different parameter layer thicknesses and since the finite element mesh is created in accordance with the parameter mesh, the initial Jacobian matrix (assuming a homogeneous ground as a starting model) cannot be precalculated and stored as a look-up file. Each parametrization scheme will result in a different initial Jacobian matrix. Thus, despite the use of the QN technique the routine for calculating the Jacobian is included in the scheme. This also gives the flexibility of starting the inversion with any desirable initial model and, most importantly, the option to chose the "traditional" slower inversion which involves full Jacobian calculations (abandoning the QN technique). This option proved most useful since a direct evaluation/comparison of the performance of the QN was possible. In Figure 6.19 a simplified flow-chart of the algorithm is depicted. A C program called 2DINVS was written in order to apply the described algorithm. The User's Manual for this program is presented in Appendix A.

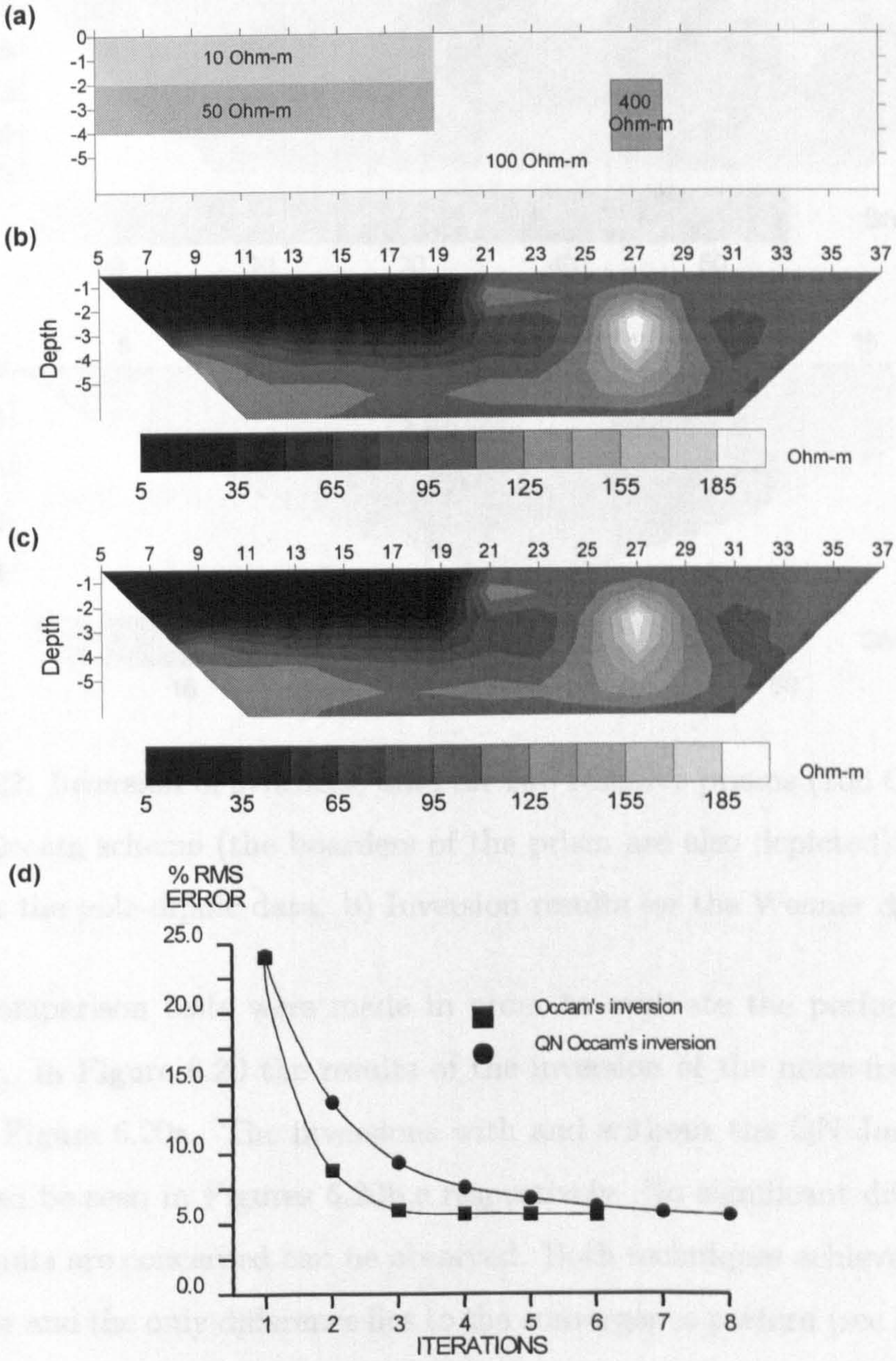


Figure 6.21: Inversion of synthetic dipole-dipole data (5% added noise, 20 electrodes, n=6): a) the model used to produce the data. b) Inversion results using the QN Occam method. c) Inversion results using the Occam method. d) The convergence of the two techniques.

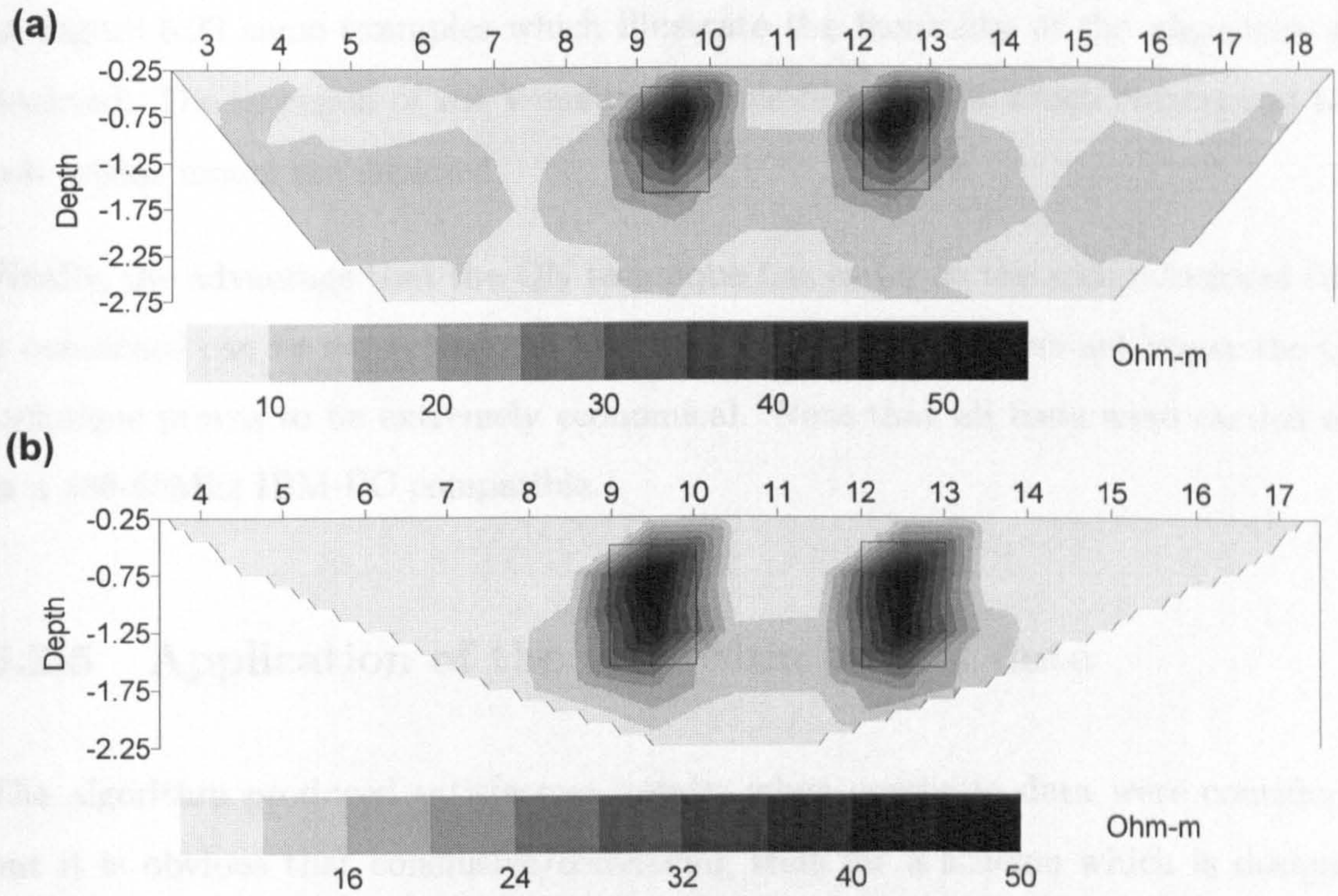


Figure 6.22: Inversion of synthetic data for two resistive prisms (100 Ohm-m) using the QN Occam scheme (the boarders of the prism are also depicted): a) Inversion results for the pole-dipole data. b) Inversion results for the Wenner data.

Several comparison tests were made in order to evaluate the performance of the algorithm. In Figure 6.20 the results of the inversion of the noise-free data of the model of Figure 6.20a. The inversions with and without the QN Jacobian matrix update can be seen in Figures 6.20b,c respectively. No significant difference as far as the results are concerned can be observed. Both techniques achieved a similar % RMS error and the only difference lies to the convergence pattern (see Figure 6.20d). The QN technique took an extra iteration (superlinear convergence as opposed to the quadratic of the "traditional" technique).

Similar results were also obtained from the inversion of the dipole-dipole data (5% random noise) from the model of Figure 6.21a. The inversions with and without the QN Jacobian matrix update can be seen in Figures 6.21b,c respectively. The convergence pattern of both techniques is depicted in Figure 6.21d.

In Figure 6.22 some examples which illustrate the flexibility of the algorithm are depicted: The inversion of the Wenner and pole-dipole data which correspond to a two prisms model are depicted.

Finally, the advantage that the QN technique has as far as the computational time is concerned can be seen *clearly in Figure 6.23. As the data sets get larger the QN* technique proves to be extremely economical. Note that all *tests were carried out* in a 486-66Mhz IBM-PC compatible.

6.2.5 Application of the algorithm to real data

The algorithm produced satisfactory results when synthetic data were considered but it is obvious that conclusive/convincing tests for a scheme which is designed for field data interpretation can only be made with real data. Furthermore, these real data sets should be from sites where there is a good knowledge of the existing targets in order to check/verify the inversion results[22] Note that most of the real-data presented here have already been discussed in the previous chapter (section 5.2.4) so only a brief description of the survey details is given.

Drain (University of York)

The data set which was obtained over the drain at the courtyard of the Department of Electronics at the University of York was inverted using the QN Occam inversion. The position of the drain in relation to the measured section is depicted in Figure 6.24a). The pseudosection of the dipole-dipole data set can be seen in Figure 6.24b (24 electrodes, dipole=60cm, n_{max}=8, 137 measurements).

The inversion after 6 iterations (7.2% RMS error) can be seen in Figure 6.24c. Despite the 3-D geometry of the target the inverted image delineates the limits of the drain quite accurately - only the top of the drain is somewhat misplaced - and

[22]In that sense data from archaeological sites are ideal since they often involve targets with well-defined shape, which can be surveyed repeatedly.

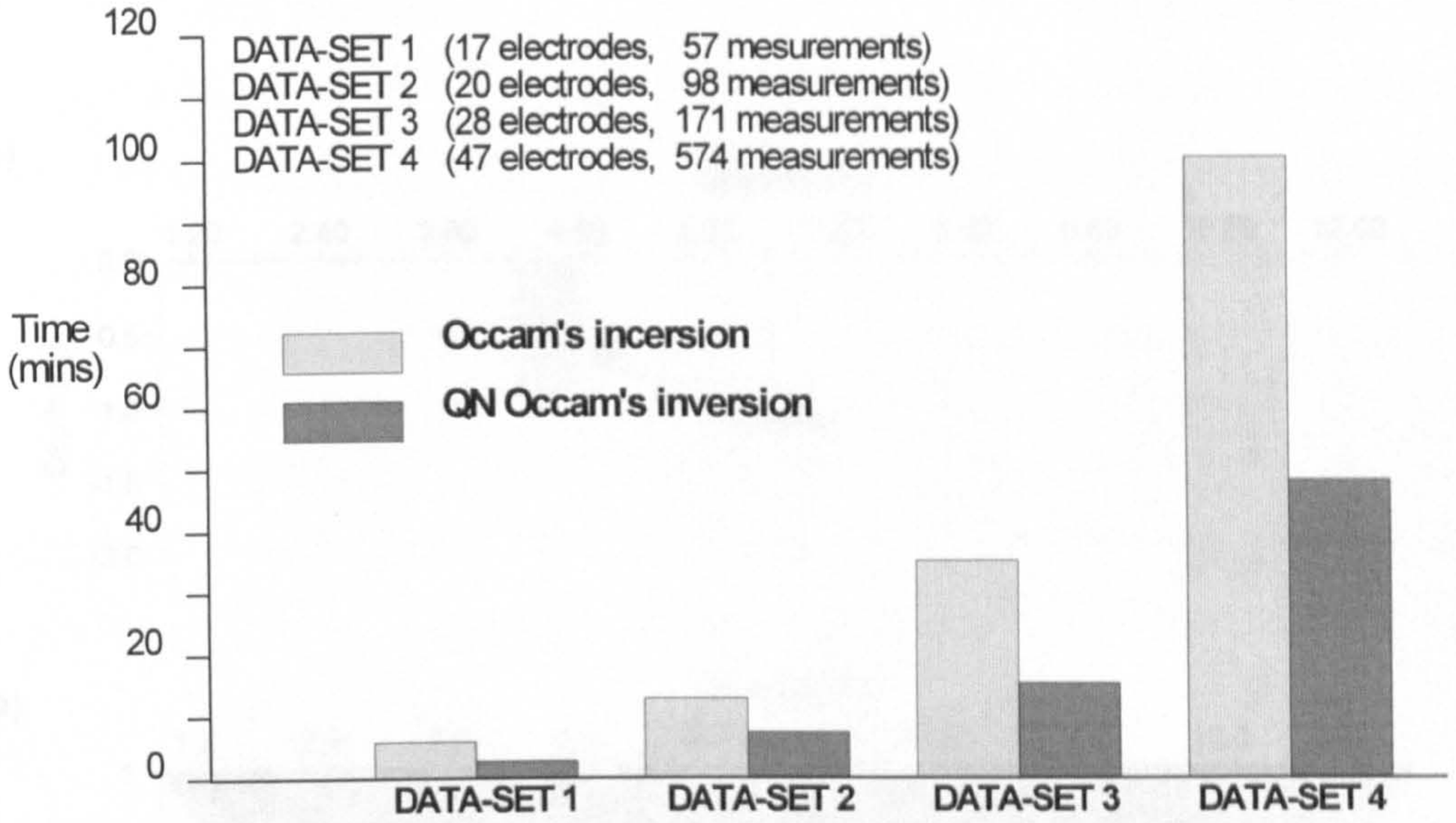


Figure 6.23: Comparison of the total run-time for 4 data sets between the QN Occam and Occam inversion schemes (all tests were performed in a 486-66Mhz IMB-PC compatible).

no major artifacts appear.

Foundation walls (Guest hall, Fountains Abbey (N. Yorkshire)

The dipole-dipole data set which was obtained over the foundations of the guest hall in Fountains Abbey (see Figures 2.13, 6.25) was inverted using the QN Occam scheme. The pseudosection of the data set (65 electrodes, dipole=50cm, n_{max}=7, 413 measurements) is depicted in Figure 6.25b.

In Figure 6.25a the dipole-dipole profile for n=1 is shown. This profile (with a separation of 0.5m) is what one would have collected from a typical archaeological geophysics profiling survey. The possible interpretation of this profile is also depicted in Figure 6.25a. The effect of the walls is clearly identified. Further, at the right-hand side of the profile (between 25 and 27 metres), a pronounced high resistive anomaly which could be interpreted as a drain is shown. Note that the full resistivity profile map (twin-probe data in Figure 2.13) justifies this interpretation.

The QN Occam inversion results (7 iterations, 2.5% RMS error) are depicted in

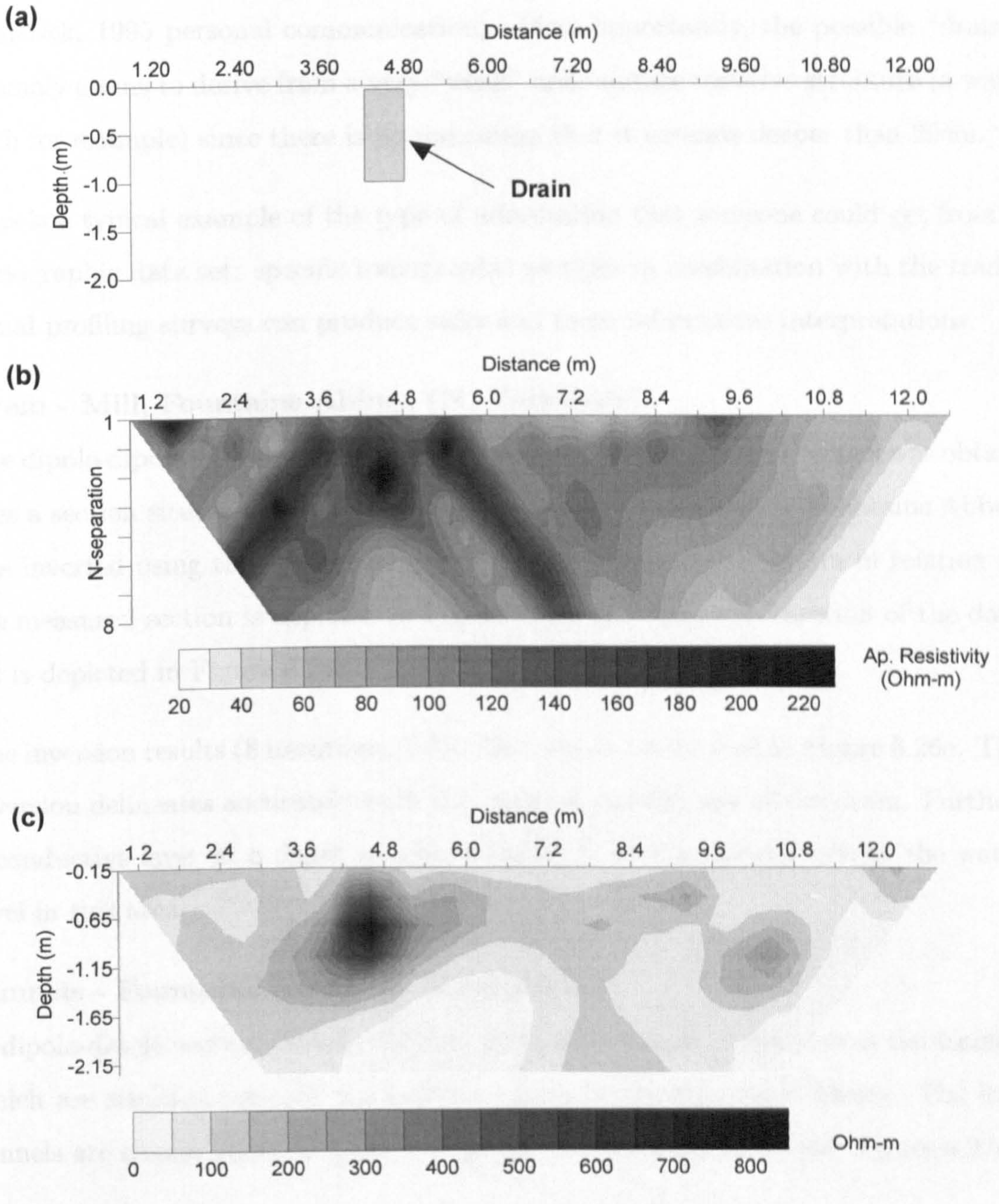


Figure 6.24: Reconstruction of the dipole-dipole data measured over a drain (University of York): a) the exact location of the drain in relation to the measured section, b) the measured data set in a pseudosection form, c) reconstruction using the QN Occam algorithm (6 iterations, 7.2% RMS error).

Figure 6.25)c. The foundation walls can be clearly seen: the lateral extent of the structures is realistic and their depth is consistent with the information about the site (Emerick, 1995 personal communication). Most importantly, the possible "drain" anomaly seems to derive from a very "weak" near-surface resistive structure (a walk path for example) since there is no indication that it extends deeper than 20cm.

This is a typical example of the type of information that someone could get from a tomographic data set: specific tomographic sections in combination with the traditional profiling surveys can produce safer and more informative interpretations.

Drain - Mill, Fountains Abbey, (N. Yorkshire)

The dipole-dipole data set (24 electrodes 1m apart, maximum n=7) which was obtain over a section situated orthogonally to a drain at the Mill area of Fountains Abbey was inverted using the QN Ocam scheme. The position of the drain in relation to the measured section is depicted in Figure 6.26a and the pseudosection of the data set is depicted in Figure 6.26b.

The inversion results (8 iterations, 1.3% RMS error) can be seen in Figure 6.26c. The inversion delineates accurately both the position and the size of the drain. Further, a conductive layer at a depth of approximately 2 metres corresponds to the water level in this area.

Tunnels - Fountains Abbey (N. Yorkshire)

A dipole-dipole section (South-to-North direction) was measured across the tunnels which are situated beneath the Infirmary area at the Fountains Abbey. The four tunnels are clearly visible and their exact dimensions are known (see Figure 6.27a). The electrode separation was 1m (37 electrodes) and the maximum n separation was 7 dipoles. The measured section does not fully "cover" the northernmost tunnel (due to obstacles). The data is presented in a pseudosection form in Figure 6.27b. It is once again clear that for a relatively complicated structure - such as this one - the pseudosection image is "uninformative" and misleading.

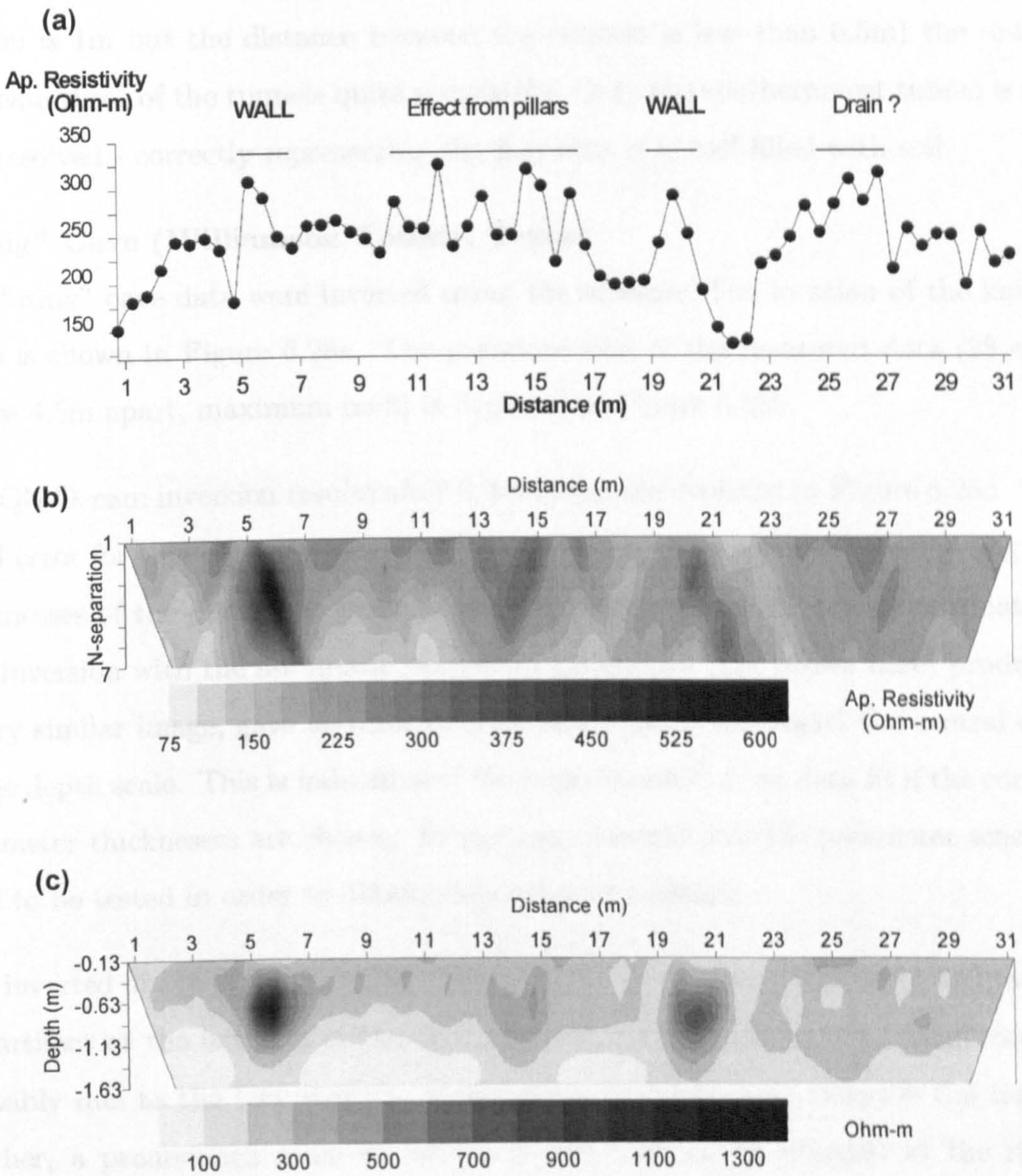


Figure 6.25: Reconstruction of the dipole-dipole data measured over the area of the guest hall at Fountains Abbey (N. Yorkshire): a) the dipole-dipole profile for n=1 (spacing =0.5m), b) the measured data set in a pseudosection form, c) reconstruction using the generalized BP algorithm (7 iterations, 2.5% RMS error).

The QN Occam inversion results (9 iterations, 3.4% RMS error) are depicted in Figure 6.27c. Despite the low lateral resolution of the survey (the electrode separation is 1m but the distance between the tunnels is less than 0.5m) the results delineate three of the tunnels quite accurately. Only the southernmost tunnel is not well resolved - correctly representing the fact that it is half-filled with soil.

"Sting" Cave (Williamson County, Texas)

The "sting" cave data were inverted using the scheme. The location of the known caves is shown in Figure 6.28a. The pseudosection of the measured data (28 electrodes 4.5m apart, maximum n=8) is depicted in Figure 6.28b.

The QN Occam inversion results after 9 iterations are depicted in Figure 6.28c. The RMS error for this inversion is 2.9%. This inversion was produced by adjusting the thicknesses of the parameter layers according to the accurate *a priori* information. The inversion with the automatic parameter generation (not shown here) produced a very similar image, gave an error of 5.4% and slightly misplaced the central cave in the depth scale. This is indicative of the improvement of the data fit if the correct parameter thicknesses are chosen. In any case, several possible parameter schemes have to be tested in order to obtain the optimum solution.

The inverted image of Figure 6.28c delineates the two known caves fairly accurately. An artifact at the left side of the anomaly that corresponds to the "sting" cave is probably due to the fact that the measurements do not fully describe the target. Further, a pronounced resistive feature is now seen to be situated at the right-hand side of the section (centre at x=98m, z=13m). Judging by the accuracy of the reconstruction of the known caves we have every reason to believe that this is another cave.

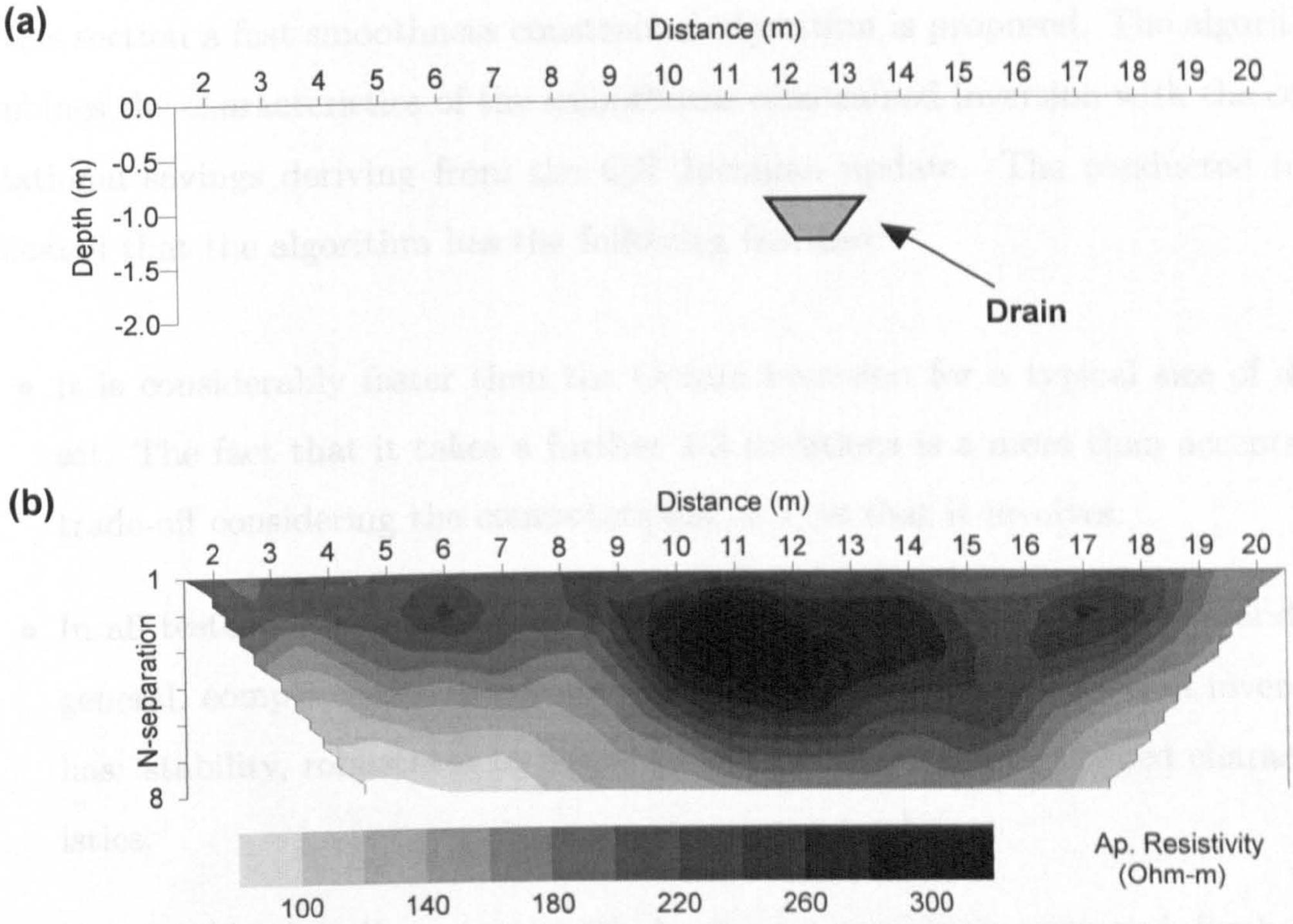


Figure 6.26: Reconstruction of the dipole-dipole data measured over a drain at the Fountains Abbey (N. Yorkshire): a) the exact location of the drain in relation to the measured section, b) the measured data set in a pseudosection form, c) reconstruction using the QN Occam algorithm (8 iterations, 1.3% RMS error).

6.2.6 Conclusions

In this section a fast smoothness constrained algorithm is proposed. The algorithm combines the characteristics of the smoothness constrained inversion with the computational savings deriving from the QN Jacobian update. The conducted tests indicated that the algorithm has the following features:

- It is considerably faster than the Occam inversion for a typical size of data set. The fact that it takes a further 1-3 iterations is a more than acceptable trade-off considering the computational savings that it involves.

- In all tested cases it produced results similar to the Occam inversion and, in general, comprises all the advantages (and limitations) that Occam inversion has: stability, robustness to noise, and inversion with user defined characteristics.

- It is flexible since it can cope with any known resistivity array and, further, it can readily cope with "unconventional" measuring schemes.

- Extra flexibility is achieved by allowing the incorporation of variable smoothness and, most importantly, variable parametrization.

- For all of the tested cases with real data the algorithm produced reasonably good results which do not suffer from algorithm and/or noise related artifacts.

Overall, the algorithm proved to be a reliable and useful tool for routine data interpretation.

6.3 Chapter overview

In this Chapter non-linear inversion techniques applied to the 2-D reconstruction of earth-resistivity data were presented.

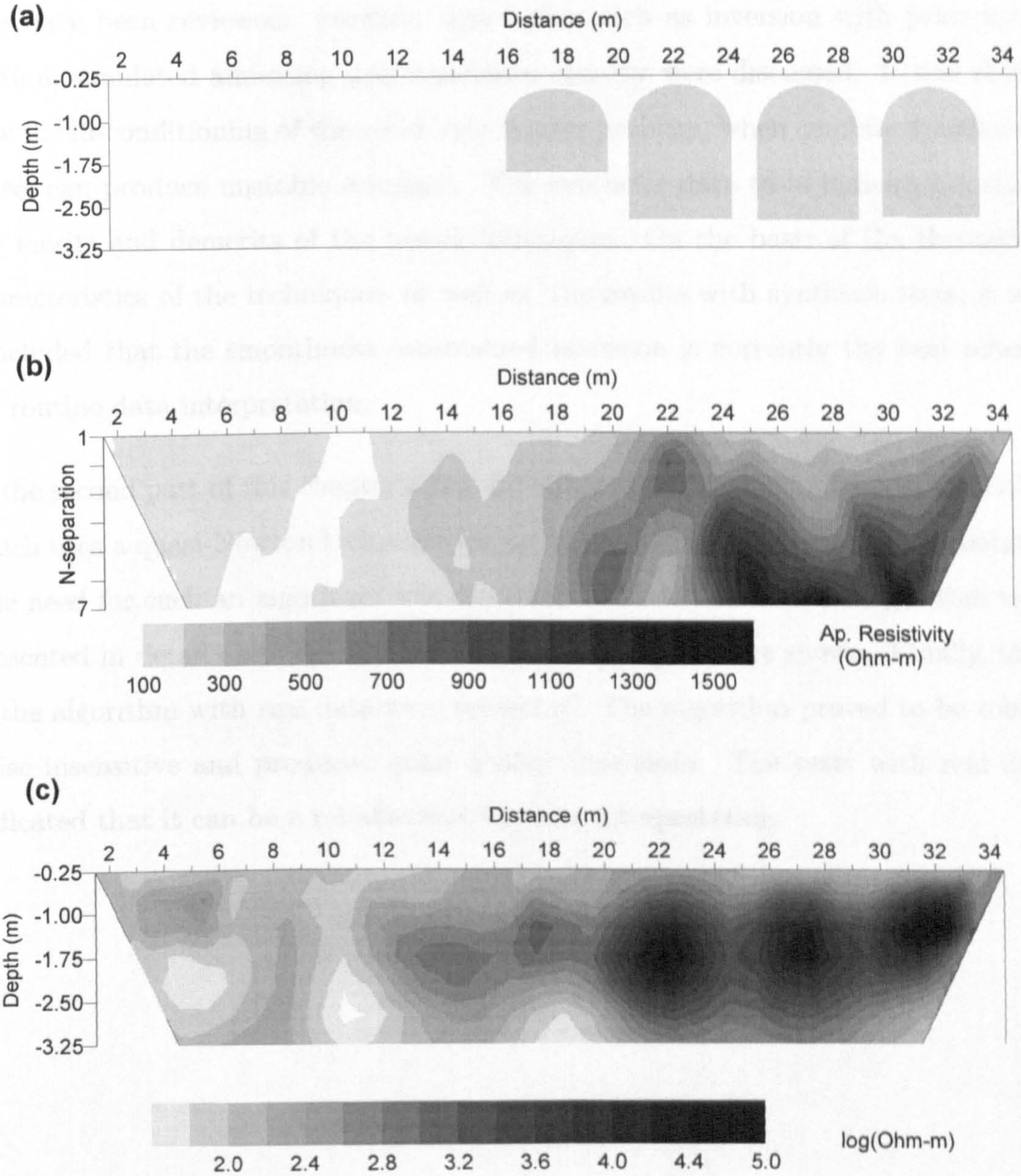


Figure 6.27: Reconstruction of the dipole-dipole data measured over the tunnels at the Infirmary of Fountains Abbey (N. Yorkshire): a) a sketch of the tunnels b) the measured data set in a pseudosection form, c) reconstruction using the QN Occam algorithm (9 iterations, 3.4% RMS error).

In the first part of this chapter, widely used techniques such as the non-linear least-squares method, Marquadt's method, and smoothness constrained (Occam) inversion have been reviewed. Further, approaches such as inversion with prior information, simulated annealing and maximum entropy were discussed. It was shown how the ill-conditioning of the resistivity inverse problem, when combined with data noise, can produce unstable solutions. The synthetic data tests indicated some of the merits and demerits of the tested techniques. On the basis of the theoretical characteristics of the techniques, as well as, the results with synthetic tests, it was concluded that the smoothness constrained inversion is currently the best scheme for routine data interpretation.

In the second part of this chapter a fast smoothness constrained inversion algorithm which uses a quasi-Newton technique for updating the Jacobian matrix is proposed. The need for such an algorithm was discussed. The features of the algorithm were presented in detail and comparisons to other techniques were shown. Finally, tests of the algorithm with real data were presented. The algorithm proved to be robust noise insensitive and produced good quality inversions. The tests with real data indicated that it can be a reliable tool for data interpretation.

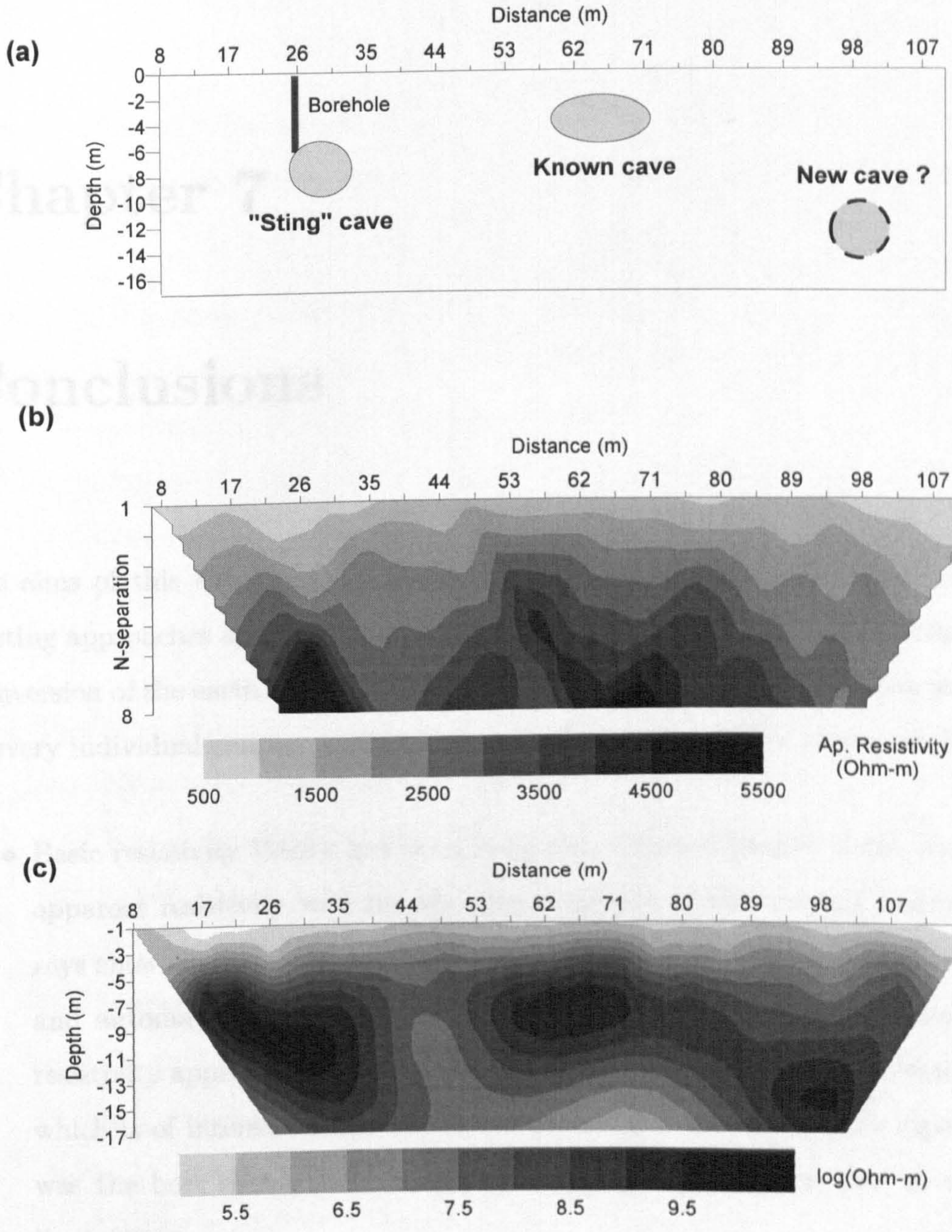


Figure 6.28: Reconstruction of the dipole-dipole data measured over caves ("Sting" Cave, Williamson County, Texas): a) the exact location of the known caves in relation to the measured section, b) the measured data set in a pseudosection form, c) reconstruction using the QN Occam scheme (9 iterations, 2.9% RMS error).

Chapter 7

Conclusions

The aims of this thesis, as described in detail in the introduction, was to study existing approaches and propose new improved algorithms for the modelling and 2-D inversion of the earth resistivity data. The conclusions drawn have been presented in every individual chapter so that only a brief overview will be presented here.

- Basic resistivity theory has been reviewed. The artificiality of the concept of apparent resistivity was discussed and a review of the existing resistivity arrays showed some of their merits and limitations. Resistivity instrumentation and automated systems were discussed and the practical application of the resistivity approach was explained. For the combined sounding profiling mode which is of interest in this thesis, it was decided that the dipole-dipole array was the best choice in view of its merits and normal hardware and survey limitations.

- The forward resistivity modelling approaches were reviewed. In view of the project's targets, the finite element method was found to be ideal for the forward resistivity modelling for the following reasons:

  a) It can incorporate any resistivity distribution.

  b) It can be used to produce the Jacobian matrix.

c) It can cope with any resistivity array.

d) it can readily be used to take terrain topography into account.

- A full description of the 2.5-D FEM scheme which was developed for this project was presented. Several comparative tests verified the accuracy of the scheme.

- A FEM scheme was proposed for modelling arrays which are parallel to the strike direction. The scheme takes full advantage of the 3-D variability of the potential and allows the modelling of arrays such as the square configuration which previously could only be modelled with full 3-D schemes. The accuracy of the scheme was tested. The results indicated that the square array and the resultant AIR can be used to delineate the exact edges of isolated targets in the profiling resistivity mode.

- The FEM scheme was used to model the effect of terrain topography on several resistivity arrays. This was necessary since the previous studies dealt only with the dipole-dipole array. The results indicated that all arrays are significantly affected by the terrain topography and that the FEM scheme can be readily used to take the terrain topography effect into consideration. Other conclusions are:

  a) All arrays produce significant, and potentially misleading, artificial errors within the data for slope angles larger than 10 degrees (if the slope extend is larger than the array spacing).

  b) The pole-pole array is more sensitive to the vertical slope topography for standard profiling applications while, conversely, the square-array is the least sensitive.

  c) For full 2-D data sets the dipole-dipole and pole-pole arrays produced anomalies of similar patterns. The Wenner array produced anomalies with opposite patterns. The pole-dipole array, in general, produced anomalies with asymmetric patterns.

d) For 2-D surveys, as far as the amplitude is concerned, dipole-dipole and pole-dipole arrays produced the highest and lowest errors when compared with pole-pole and Wenner arrays. However, direct quantitative comparisons cannot be made since the sensitivity of each array is different.

- It was showed that the Jacobian matrix plays a significant role within the inversion procedure. Ways for calculating the Jacobian (sensitivity technique, adjoint equation technique, perturbation technique) were presented within the FEM procedure. The equations for applying these techniques were developed explicitly. The adjoint equation technique was found to be accurate and quite economical in computing time. The adjoint equation technique was used to calculate the sensitivity response of common resistivity arrays. The results indicate that:

  a) The Jacobian matrix can be used to get a qualitative analysis of the response of the arrays over known targets.

  b) It was shown that unusual apparent resistivity responses can be explained adequately by the sensitivity of the arrays.

  c) The Jacobian matrix can be used as a means for designing optimum surveys or achieving optimum parametrization.

- Approximate resistivity techniques have been reviewed. The theoretical analysis of their features and results from synthetic data indicated that:

  a) The pseudosection technique is based on crude assumptions but it is simple in its implementation. It produced results which, in most cases, suffer from major artifacts which are associated with the varying sensitivity of the arrays.

  b) The associated Barker's method suffers from the same limitations. Despite its iterative character, it has no self-correcting mechanism for adjusting the spatial resistivity distribution. It is able to produce reliable reconstructions only when the pseudosection assumptions are valid.

  c) The user-forward modelling interactive technique is heavily based on the operator's expertise and is subject to biased implementation. Further, it can-

not cope with the increased amount of data produced by automated measuring systems.

d) The simple back-projection algorithms (such as Bristow's method and Powell's method) are quite crude but easily applicable since there is no need to calculate the Jacobian matrix. The equipotential constraint can be applied to only a limited number of arrays. When it works, it guarantees that BP will take place at the most sensitive regions of the subsurface. The associated Noel's method is more sophisticated (it makes use of the Jacobian matrix) but still suffers from reconstruction associated artifacts.

e) The BP technique proposed by Shima cannot be used directly for reconstructing surface resistivity data since it produces negative resistivity values.

- On the basis of the theoretical characteristic of the approximate reconstruction techniques tested in this work a generalized back-projection algorithm was proposed. The algorithm :

a) Can include many of the existing approximate algorithms.

b) Can cope with any type of measuring scheme.

c) It can be iterative, and thus produce both qualitative and quantitative information of the resistivity distribution.

Those are achieved by recognizing the major importance of the Jacobian matrix within the approximate reconstruction procedure.

The tests with synthetic data indicated that the algorithm:

a) produces reconstructions which do not suffer from major artifacts.

b) produces reasonably accurate qualitative and quantitative reconstructions.

c) it is relatively noise-insensitive.

The tests with real data indicated that the algorithm:

a) produced results that are generally in good agreement with the known targets.

b) in some case-studies (of a generally complicated nature) failed to delineate

the depth of the targets successfully. This problems are due to its approximate nature. For the same reason the RMS errors were not particularly low.

- Several non-linear techniques used in the solution of the 2-D resistivity inverse problem were presented. It was shown that ill-conditioning combined with data-errors generates unstable solutions. As far as the tested techniques are concerned the following can be said:

  a) Marquadt's method produced very good results when noise-free data were tested, but in the case of noisy data spurious noise-related artifacts appeared. More generally, the problem with Marquadt's method is that it is based on a mathematical "trick" which produces stability but cannot guarantee physically realistic solutions.

  b) Smoothness constrained inversion produces a solution which has the general properties that the user has *a priori* selected. The tests with noisy data indicated that the smoothness constraint is successful in avoiding noise related artifacts. Furthermore, the technique proved particularly robust and although it did not produce the "exact" solutions the results were quite close to reality.

  c) Inversion with *a priori* information is theoretically quite powerful: apart from general constraints such as smoothness, prior information regarding the spatial variation of resistivity can be included. But when this type of prior information is not reliable, inversion with *a priori* information can lead to biased interpretations and generally adds extra worries into the interpretation procedure.

  d) The maximum entropy technique looked quite promising since (due to its inherent smoothness) it can cope well with noisy data. More tests with synthetic and real data are needed in order to evaluate its performance.

  e) The simulated annealing technique needs further testing. The general characteristics of the method (global searching and avoiding being trapped in local minima) are quite appealing, but currently it's applicability is limited by the enormous computational loads that it involves.

In conclusion, it is believed that currently the smoothness constrained inversion is the most preferable technique for practical data interpretation. Although far from perfect, it comprises numerous advantages: **a)** it is physically reasonable, **b)** it is robust, **c)** it can cope well with noise and does not produce artifacts, **d)** it produces solutions which have properties that the interpreter has chosen and are not a product of an arbitrary initial choice.

- A fast smoothness constrained algorithm was proposed. The algorithm combines the characteristics of the smoothness constrained inversion with the overall computational savings deriving from the QN Jacobian update. The conducted tests indicated that the algorithm has the following features:

  a) It is considerably faster than the Occam inversion for a typical size of data set. The fact that it takes a further 1-3 iterations is a more than acceptable trade-off considering the computational savings that it involves.

  b) In all cases tested it produced results similar to the Occam inversion and, in general, comprises all the advantages (and limitations) that Occam inversion has: stability, robustness to noise, and inversion with user defined characteristics.

  c) It is flexible since it can cope with any known resistivity array and, further, it can readily cope with "unconventional" measuring schemes.

  d) Extra flexibility is achieved by allowing the incorporation of variable smoothness and, most importantly, variable parametrization.

  e) For all of the tested cases with real data the algorithm produced reasonably good results which did not suffer from algorithm and/or noise related artifacts.

  Overall, the algorithm proved to be a reliable and useful tool for routine data interpretation.

Some more general conclusions that can be drawn from the study are:

- The results with the real data sets illustrated the great potential that inversion techniques have. It was shown that inversion can delineate even complicated

structures and produce valuable information which can significantly increase the knowledge of a site.

- On the other hand, it was illustrated that the inversion techniques have limitations and can produce artifacts (especially the approximate techniques). In this context, the user should be aware of the possible limitations of the techniques in order to avoid erroneous interpretations.

- For most of the inversions presented in this work a remarkably good fit between the modelled and real data was achieved. This is not only due to the fact that the presented inversion schemes are relatively efficient but also due to other factors which are equally important for obtaining good quality results:
  - Most of the presented data sets are from sites which are geologically "clean" (i.e. Fountains Abbey): there are a limited number of near-surface inhomogeneities which could mask the effect of the targets (geological "noise").
  - The majority of the tested targets satisfy (to a certain degree) the assumptions of the 2.5-D modelling ("infinite" extent along the strike (y) direction).
  - Most importantly, the quality of the collected data was particularly good and this is a consequence of the resistivity instrumentation. The Sting/Swift system (AGI Inc.), with which most of the data was collected, has automatic noise-control features and produced reliable dipole-dipole measurements (recall that the dipole-dipole array has a low signal-to-noise ration).

- Direct comparisons between the approximate and accurate techniques cannot be made. The generalized back-projection algorithm produced results which are inferior to the fast QN Occam scheme - a consequence of the simplifying assumptions of the approximate schemes. Further, the back-projection algorithm proved to be slightly slower than the QN algorithm for the typical sized data sets presented in this work. On the other hand, preliminary results indicate that the BP algorithm is much faster when particularly large data sets are considered since it avoids the matrix inversion procedure. In that sense,

the generalized BP algorithm is best suited to the preliminary interpretation of sizeable data sets.

7.1 Future Work and Lines of Research

This work is by no means a *complete* investigation of the tested algorithms. Only some particulars features of the presented algorithms were studied having in mind mainly the practical application of the schemes. Therefore, there is much left to be done.

Further study of the theoretical properties of the presented algorithms is necessary. We believe that this is essential since numerical tests, on their own, can be indicative of the performance of an algorithm but cannot be conclusive. A combined theoretical and practical study is the only way to find out the functional limitations and advantages of the various techniques.

Moreover, since all the algorithms are designed for addressing a real problem, further tests with real data are needed. This data should be measured over calibrated test-sites and should describe various targets with different degrees of complexity and in varying contexts.

As far as the schemes presented in this work are concerned the following can be done:

Modelling

The accuracy of the forward modelling is essential for the quality of the inversion results. With regard to the FEM modelling several modifications can be made in order to improve the accuracy: for example, the investigation of more efficient boundary conditions such as the application of "boundary elements" (Zienkiewicz and Taylor, 1989) or the incorporation of a context-adjustable mesh (automatic increase of the mesh density in areas of large property contrast).

Approximate inversion schemes

As the potential size of the resistivity data sets is increasing, the approximate algorithms will always be a useful tool for preliminary data interpretation. The generalized BP algorithm presented in this work has to be tested further with synthetic and real data for the cases of full tomographic data sets since, as was explained, we believe that this is a field where its application is more practical.

Accurate inversion schemes

With respect to the accurate inversion schemes, there is a scope for the following:

A better understanding of the effect of the different smoothness patterns in Occam inversion is necessary. Further, the technique can be used in conjunction with more efficient matrix inversion schemes such as the generalized singular value decomposition which are particularly suited for solving regularization problems (Golub and Van Loan, 1989).

Further research on the practical application of *a priori* information in the resistivity inversion is needed. Although the theoretical background of the relevant algorithms is fully developed, there is very limited practical experience in this field.

Further theoretical and practical evaluation of the Quasi-Newton techniques is necessary. Moreover, alternative QN updating formulas can be tested (see Gill and Murry (1974) for example).

Finally, the preliminary results indicate that there is a scope for further studying and developing techniques such as the maximum entropy and simulated annealing.

Increase in speed

Due to the advent of the automated systems there is a scope for further accelerating the inversion/interpretation procedure. This can be achieved either by means of hardware (more powerful computers, parallel processing etc.), efficient programming (i.e. machine code) and/or more efficient algorithms. Already, fast algorithms for inverting large matrices such as the sub-space methods (Oldenburg et al., 1993)

and conjugate gradient methods (Zhang, 1995) have been applied to the resistivity inverse problem with satisfactory results.

3-D Inversion

Most of the algorithms presented in this work can be modified readily in order to solve 3-D inversion problems. 3-D schemes are currently memory and time consuming but increases in the inversion speed (discussed above) will enable 3-D inversion to be widely used: so there is no doubt that in a few years time 3-D schemes will be the standard procedure for modelling and interpreting resistivity data.

Chapter 8

References

Abramowitz, M., and Stegun, I. (1972). Handbook of mathematical functions. Dover Publications Inc.

Aitken, M. (1974). Physics and archaeology. Clarendon Press, Oxford.

Al Hagrey, S.A. (1994). Electric study of fracture anisotropy at Flakenberg, Germany. Geophysics, **59**, 881-889.

Al'pin, L.M. (1950). The theory of dipole sounding: in A Selection of Papers on Dipole Techniques, Keller, G.V. (ed.), 1-50.

Andrews, H., and Hunt, B. (1977). Digital image restoration. Pentice-Hall Inc.

Apparao, A., Gangadhara, R., Sivarama, S., and Subrahmanya, S. (1992). Depth of detection of buried conductive targets with different electrode arrays in resistivity prospecting. Geophysical Prospecting, **40**, 749-760.

Artola, J. (1994). A study on electrical impedance tomography reconstruction algorithms. Ph.D. Thesis. University of York.

Aspinall, A., and Lynam, J.T. (1970). An induced polarization instrument for the detection of near surface features. Prospezzioni Archaeologiche, **5**, 67-75.

Aubert, M., Camus, G., and Fournier, C. (1984). Resistivity and magnetic surveys in ground-water prospecting in volcanic areas - case history Maar of Beanit, Puy de Dome, France. Geophysical prospecting, **32**, 554-563.

Bibby, H.M., and Hohmann, G.W. (1993). 3-D interpretation of multiple source bipole-dipole resistivity data using the apparent resistivity tensor. Geophysical prospecting, **41**, 697-723.

Breckom, W., and Pidock, M. (1987). Mathematical aspects of impedance imaging. Clin. Phys. Physiol. Meas. Suppl.A, **8**, 77-84.

Barber, D.C., Brown, B.H., and Freeston, I.L. (1983). Imaging spatial distributions of resistivity using applied potential tomography. Electronic Letters, **20**, 933-935.

Barber, D.C., and Brown, B.H. (1984). Applied potential tomography. J. Phys. E: Sci. Instrum., **17**, 723-732.

Barber, D.C., and Brown, B.H. (1987). Fast reconstruction of resistance images. Clin. Phys. Physiol. Meas. Suppl.A, **8**, 47-54.

Barker, R.D. (1981). The offset system of electrical resistivity sounding and its use with a multi-core cable. Geophysical Prospecting, **29**, 128-143.

Barker, R.D. (1989). Depth of investigation of collinear symmetrical four-electrode arrays. Geophysics, **54**, 1031-1037.

Barker, R. (1992). A simple algorithm for electrical imaging of the subsurface. First Break, **10**, (2), 53-63.

Beard, L.P., and Morgan f.D. (1991). Assessment of 2-D resistivity structures using 1-D inversion. Geophysics, **56**, 874-883.

Bernabini, M., Brizzolari, E., and Mazzola, C. (1987). Anomalies due to resistive parallelepiped bodies in resistivity profiles. Bollettino Di Geofisica Theoritica Et Applicata, **29**, 133-146.

Bonomi, E., and Lutton, J.L (1984). The N-city travelling salesman problem: sta-

tistical mechanics and the metropolis algorithm. SIAM Review, **26**, 551-568.

Box, G.E.P., and Kanemasu, H. (1972). Topics in model building, part II: on non-linear least squares. Technical Report No. 321. Department of Statistics, University of Wisconsin, Madison.

Bristow, C.M. (1966). A new graphical resistivity technique for detecting air-filled cavities. Study in Speleology, **1**, 204-227.

Broyden, C.G. (1965). A class of methods for solving nonlinear simultaneous equations. Mathematics of Computing, **19**, 577-593.

Broyden, C.G. (1970). Recent developments in solving nonlinear algebraic systems: in Numerical Methods for Nonlinear Algebraic Equations, Rabinowitz, P. (ed.), Gordon and Breach Science Publishers, 61-74.

Burden, R., and Faires, J.D. (1993). Numerical analysis. 5th edition, PWS-Kent Publishing Company, Boston.

Burnett, D.S. (1989). Finite element analysis. Addison-Wesley Publishing Co.

Butler, D.K., and Llopis, J.M. (1990). Assessment of anomalous seepage conditions: in Investigations in Geophysics no 5, Geotechnical and Environmental Geophysics vol II, S. Ward (ed.), SEG, Tulsa, 153-173.

Carpenter, E.W., and Habberjam, G.M. (1956). A tripotential method for resistivity prospecting. Geophysics, **21**, 455-469.

Clark, A. (1990). Seeing beneath the soil: prospecting methods in archaeology. Batsford, B.T. Ltd, London.

Christakos, G. (1991). Some applications of the Bayesian , maximum entropy concept in geostatistics: in Maximum Entropy and Bayesian Methods, Grandy, W.T. and Schick, K.L. (eds.), Kluwer, NL, 215-229.

Coggon, J.H. (1971). Electromagnetic and electrical modelling by the finite element method. Geophysics, **36**, 132-155.

Coggon, J.H. (1973). A comparison of IP electrode arrays. Geophysics, **38**, 737-761.

Constable, S. Parker, R., and Constable C. (1987). Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. Geophysics, **52**, 289-300.

Cook, K.L., and Van Nostrand, R.G. (1954). Interpretation of Resistivity Data Over Filled Sinks. Geophysics, **19**, 761-790.

deGroot-Hedlin, C., and Constable, S (1990). Occam's inversion to generate smooth, two-dimensional models from magnetoteluric data. Geophysics, **55**, 1613-1624.

Dahlin, T. (1993). On the automation of 2D resistivity surveying for engineering and environmental applications. Ph.D. Thesis, Lund University.

Dahlin, T., Johanssson, S., and Landin, O. (1994). Resistivity surveying for planning of infrastructure. Proceedings of SAGEEP 94, Boston Massachusetts, March 27-31, 1994, 509-528.

Daily, W., Ramirez, A., LaBrecque, D., and Barber, W. (1995). electrical resistance tomography experiments at the Oregon Graduate Institute. Journal of Applied Geophysics, **33**, 227-237.

Daniels, D., Gunton, D., and Scott, H. (1988). Introduction to subsurface radar. IEE Proceedings, **135**, Pt. F., (4), 278-320.

Das, U.C., and Parasnis, D.S. (1987). Resistivity and Induced Polarization Responses of Arbitrarily Shaped 3-D Bodies in a Two-Layered Earth. Geophysical Prospecting, **35**, 98-109.

Dey, A., Meyer, W. Morrison, F., and Dolan, W (1975). Electric field response of 2-D inhomogeneities to unipolar and bipolar electrode configurations. Geophysics, **40**, 630-640.

Dey, A., and Morrison, H.F. (1979a). Resistivity modelling for arbitrarily two-dimensional structures. Geophysical Prospecting, **27**, 106-136.

Dey, A., and Morrison, H.F. (1979b). Resistivity modelling for arbitrarily three-dimensional structures. Geophysics, **49**, 753-780.

Dines, K., and Lytle, J. (1979). Computerized geophysical tomography. Proceedings of the IEEE, **67**, 1065-1073.

Dittmer, J., and Szymanski, J. (1992). The forward modelling of resistive two-dimensional features using the finite element method. Monograph on Geophysical Data Inversion in Archaeological Site Investigation, Vieweg Verlag (in press).

Dittmer, J.K. (1994). The stochastic inversion of magnetic and resistivity data with the simulated annealing algorithm. Ph.D Thesis, University of York.

Dittmer, J.K., and Szymanski, J.E. (1995). The stochastic inversion of magnetic and resistivity data with the simulated annealing algorithm. Geophysical Prospecting , **43**, 397-416.

Edwards, L.S. (1977). A Modified Pseudosection for Resistivity and IP. Geophysics, **42**, 1020-1036.

Ellis, R.G., and Oldenburg D.W. (1994). Applied geophysical inversion. Geophys. J. Int., **116**, 5-11.

Eyuboglu, B.M., Brown, B.H., and Barber, D.C. (1989). In vivo imaging of cardiac related impedance changes. IEEE Engineering in Medicine and Biology Magazine, March 1989, 39-45.

Fletcher, R. (1987). Practical methods of optimization. 2nd edition, Wiley Ltd.

Fox, R.C., Hohmann, G.W., Killpack, T.J., and Rijo, L. (1980). Topographic effects in resistivity and induced polarization surveys. Geophysics, **45**, 75-93.

Fristiani, A.L., Molinari, G., and Viviani, A. (1980). Introduction: in Finite Elements in Electrical and Magnetic Field Problems, , Chari, M. and Silvester, P.(eds). Cambridge University Press, 1-10.

Fox, R., Hohmann, G., Killpack, T., and Rijo, L. (1980). Topographic effects in

resistivity and induced polarization Surveys. Geophysics, **45**, 75-93.

Furness, P. (1992). An Integral equation for the geoelectric response of thin resistive bodies. Geophysical Prospecting, **40**, 701-720.

Gilbert, P. (1972). Iterative methods for the reconstruction of 3-D objects from projections. J. Theoret. Biol., **36**, 132-155.

Gill, P.E., and Murray, W. (1974). Numerical methods for constrained optimization. Academic Press.

Golub, H.G., and Van Loan F.C. (1989). Matrix computations. 2nd Edition, The John Hopkins University Press.

Gordon, R. (1974). A Tutorial on ART (Algebraic Reconstruction Technique). IEEE Trans. Nuc. Sc., 21, 78-92.

Gordon R., Herman, G., and Johnson, S. (1985). Image reconstruction from projections. Sci. Amer., 233, 56-58.

Griffiths D., Turnbull, J., and Olayinka, A. (1990). Two-dimensional resistivity mapping with a computer-controlled Array. First Break, **8** (4), 121-129.

Griffiths, D., and Barker, R. (1993). Two dimensional resistivity imaging and modelling in areas of complex geology. Journal of Applied Geophysics, **19**, 211-226.

Gull, S.F., and Skilling, J. (1984a). The maximum entropy method: in Indirect Imaging, Rosenkrantz, R (ed.), Cambridge University Press.

Gull, S.F., and Skilling, J. (1984b). Maximum entropy method in image processing. IEE Proceedings, **131** (F), 646-690.

Habberjam, G.M., (1975). Apparent resistivity, anisotropy and strike measurements. Geophysical Prospecting, **23**, 211-247.

Hanson, K. (1987). Bayesian and related methods in image reconstruction from Incomplete Data: in Image Recovery: Theory and application, Stark, H. (edt).

Academic Press, Inc. 79-125.

Herman, G., Lent, A., and Rowland, S. (1973). ART: mathematics and applications. J theor. Biol., **42**, 1-32.

Herman, G. (1980). Image reconstruction from projections: The fundamentals of computed tomography. Academic Press Inc.

Hesse, A., Jolivet, A., and Tabbagh, A. (1986). New prospects in shallow depth electrical surveying for archaeological and pedological Applications. Geophysics, **51**, 585-594.

Hobbs, B.A., and Reading, A.M. (1994). Shallow fault location in coal measures using offset Wenner resistivity profiling. Geophysical Prospecting, **42**, 343-356.

Hohmann, G. (1988). Numerical modelling for EM methods of geophysics: in Electromagnetic Methods in Applied Geophysics, Nabighian, M.N. (ed.) vol 1, Ch. 5. SEG, Tulsa, 314-364.

Hohmann, G., and Raiche, A. (1988). Inversion of controlled source EM data: in Electromagnetic Methods in Applied Geophysics, Nabighian , M.N. (ed.) vol 1, Ch. 8. SEG, Tulsa, 469-504.

Holcombe, J., and Jiracek, G. (1984). 3-D terrain corrections in resistivity surveys. Geophysics, **49**, 439-452.

Ider, Z., Gencer, N., Atalar, E., and Tosun, H. (1990). Electrical impedance tomography of translational uniform cylindrical objects with general Cross-sectional boundaries. IEEE Trans. Med. Img., **9**, 49-59.

Ihnman, J., Ryu, J., and Ward, S. (1973). Resistivity inversion. Geophysics, **38**, 1088-1108.

Jackson, D., and Matsu'ura, M. (1985). A bayesian approach to nonlinear inversion. J. Geophys. Research, **90** (1), 581-591.

Jupp, D., and Vozoff, K. (1975). Stable iterative methods for the inversion of

geophysical data. Geophys. J. R. astr. Soc., **42**, 957-976.

Keller, G.V., and Frischknecht, F.C. (1966). Electrical methods in geophysical prospecting. Pergamon Press, New York, NY.

Khon, R., and Vogelius, M. (1984). Determining conductivity by boundary measurements. Commun. Pure and Appl. Math., **37**, 289-298.

Kim, Y., Webster, J., and Tompkins, W. (1983). Electrical impedance imaging of the thorax. J. Microwave Power, **18**, 245-257.

Kim, Y., and Woo, H. (1987). A Prototype system and reconstruction algorithms for electrical impedance technique in medical body imaging. Clin. Phys. Physiol. Meas., **8** (Suppl. A), 63-70.

Koefoed, O. (1979). Geosounding principles, 1, Resistivity Sounding Measurements. Elsevier, Amsterdam.

Kohlbeck, F., Szarka, L., Steiner, T. Hollo, L., and Muller, I. (1993). Lake-bottom geoelectric and water born VLF measurements on the lake Fertu (Neusiedlersee) Expanded abstracts of the 55th meeting and technical exhibition of the EAEG: Stavanger, Norway, 7-11 June, 1994, D051.

Korte, C.J. (1993). Sensitivity coefficient method for the reconstruction of electrical impedance tomography. Clin. Phys. Physiol. Meas., **10**, 275-281.

Korte, C.J. (1993). Studies of image reconstruction methods for electrical impedance tomography. Ph.D. Thesis. University of Newcastle-upon-Tyne.

Kreyszing, E. (1992). Advanced engineering mathematics. John Wiley and Sons.

Kunetz, G. (1966). Principles of direct current resistivity prospecting. Gebruder Borntaeger, Berlin.

Lanczos, C. (1960). Linear differential operators. D. Van Nostrad Company Ltd.

Lawson, C., and Hanson, R. (1974). Solving least squares problems. Pentice-Hall.

Lee, T. (1975). An integral equation and its solution for some two and three-dimensional problems in resistivity and induced polarization. Geoph. J. Roy. Astr. Soc., **45**, 1-95.

Li, Y., and Oldenburg, W. 1992). Approximate inverse mapping in DC resistivity problems. Geophys. J. Int., **109**, 343-362.

Lines, L.R., and Treitel S. (1984). Tutorial: a review of least-squares inversion and its application to geophysical problems. Geophysical Prospecting, **32**, 159-186.

Lowry, T., Allen M.B, and Shive, P. (1989). Singularity removal: a refinement of resistivity modeling techniques. Geophysics, **54**, 766-774.

Lowry, T., and Shive, P. (1990). An evaluation of Bristow's method for the detection of subsurface cavities. Geophysics, **55**, 514-520.

McGillivray, P., and Oldenburg, D. (1990). Methods for calculating Frechet derivatives and sensitivities for the Non-linear inverse problem: A comparative study. Geophysical Prospecting, **38**, 499-524.

McNeil, J.D. (1980). Electrical conductivity of soils and rocks. Technical Report, Geonics Ltd, 5-22.

Marquadt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Indust. Appl. Math., **11**, 431-441.

Marron, M. (1982). Numerical Analysis; a Practical Approach. Macmillan, N.York.

Molano, C.E., Salamanca, M.M., and van Overmeeren R.A. (1990). Numerical modelling of standard and continuous vertical electrical soundings. Geophysical Prospecting **38**, 705-718.

Mufti, L. (1976). Finite-difference resistivity modelling for arbitrarily shaped two-dimensional structures. Geophysics, **41**, 62-78.

Murai, T., and Kagawa, Y. (1985). Electrical impedance computed tomography based on a finite element model. IEEE Trans. Biom. Eng., **32**, 177-184.

Noel, M., and Xu, B. (1991). Archaeological investigation by electrical resistivity tomography: a preliminary report. Geophys. J. Int., **107**, 95-102.

Noel, M., and Walker R. (1992). Imaging archaeology by electrical resistivity tomography: a preliminary study. in Archaeological sciences 89, Budd, P., Chapman, B., Jackson, C. Janaway, R. and Ottaway, B., Oxbow, 295-304.

Noel, M. (1992). Multielectrode resistivity tomography for imaging archaeology: in Geoprospection in the Archaeological Landscape, Spoerry, P. (ed.), Oxbow Monograph **18**, 89-99.

Olayinka, A., and Barker, R. (1990). Borehole siting in crystalline basement areas of Nigeria with microprocessor-Controlled resistivity traversing system. Ground Water, **28**, (2), 178-183.

Olesen, O., Henkel, H., Lile, O., Mauring, E., and Ronning, J. (1992). Geophysical investigations of the Stuorangua postglacial fault, Finnmark, N. Norway. Journal of Applied Geophysics, **29**, 95-118.

Oldenburg, D.W. McGillivray, P.R., and Ellis, R.G. (1993). Generalized subspace methods for large-scale inverse problems. Geophys. J. Int., **114**, 12-20.

Orlando, L., Piros, S., and Versino, L (1987). Location of subsurface geoelectric anomalies for archaeological work: a comparison between experimental arrays and interpretation using numerical methods. Geoexploration, **24**, 227-237.

Overmeeren, R.A., and Ritsema, I.L. (1988). Continuous vertical electrical sounding, First Break, **6**, (10), 313-324.

Patella, D. (1978). Application of geoelectric dipolar techniques to the study of an underground natural cavity of archaeological interest. Bollettino Di Geophysica Teorica et Applicata, **21**, 23-34.

Parasnis, D. (1986). Principles of applied geophysics. Chapman and Hall.

Park, S.K., and Van, G.P. (1991). Inversion of pole-pole data for 3-D resistivity

structure beneath arrays of electrodes. Geophysics, **56**, 951-960.

Powell, M.J.D. (1970). A hybrid method for nonlinear equations: in Numerical Methods for Nonlinear Algebraic Equations, Rabinowitz, P. (ed.), Gordon and Breach Science Bublishers, 61-74.

Pedersen, L.B., and Rasmunssen, T.M. (1989). Inversion of magentotelluric data: a non-linear least-squares approach. Geophysical Prospecting, **37**, 669-695.

Pelton, W., Ward,S. Halloff, P., Sill, W., and Nelson P. (1978). Mineral discrimination and removal of inductive coupling with multifrequency induced polarization. Geophysics, **43**, 588-609.

Pelton, W., Rijo, L., and Swift, J. (1978). Inversion of two-dimensional resistivity and induced polarization Data. Geophysics, **43**, 788-803.

Petrick, W., Pelton, W., and Ward, S. (1981). Ridge regression inversion applied to crust resistivity sounding data from South Africa. Geophysics, **42**, 995-1005.

Petrick, W., Sill, W., and Ward, S. (1981). Three-dimensional resistivity inversion using alpha centers. Geophysics, **46**, 1148-1162.

Powell, H.M., Barber, D.C., and Freeston, I.L. (1987). Impedance imaging using linear electrode arrays. Clin. Phys. Physiol. Meas. Suppl.A, **8**, 109-118.

Press, W., Flannery, B., Teukolski, S., and Vetterling, W. (1987). Numerical recipes: the art of scientific computing. Cambridge University Press.

Pous, J., Marcuello, A., and Queralt, P. (1987). Resistivity inversion with a priori information. Geophysical Prospecting, **35**, 590-603.

Pridmore, D., Hohmann, G., Ward, S., and Sill, W. (1981). An investigation of finite-element modelling for electrical and electromagnetic data in three dimensions. Geophysics, **46**, 1009-1024.

Queralt, P., Pous, J., and Marcuello A. (1991). 2-D resistivity modelling: an approach to arrays parallel to the strike direction. Geophysics, **56**, 941-950.

Radcliff, R., and Balanis, C. (1979) Reconstruction algorithms for geophysical applications in Noisy Environments. Proc. IEEE, **67**, (7), 1060-1064.

Rao, D. (1982) The Finite element method in engineering. Pergamon Press.

Rijo, L., (1977). Modeling of electric and electromagnetic data. Ph.D. Thesis. University of Utah.

Rijo, L., Pelton, W.H., Feitosa, E.C., and Ward, S.H. (1977). Interpretation of apparent resistivity data from the Apodi Valley Rio Grande do Norte Brazil. Geophysics, **42**, 811-822.

Robinson, E. (1985). Applied seismology: in Proceedings of the First and Second International Workshops on Maximum Entropy and Bayesian Methods in Applied Statistics, University of Wyoming, U.S.A., June 8-10, 1981, August 9-11, 1982, Smith, C. and Grandy, W. (eds.), Reidel, 211-242.

Robinson, E., and Coruh, C. (1988). Basic exploration geophysics. John Wiley and Sons.

Rodgers, R.B., and Kean, W.F. (1980). Monitoring ground-water contamination at a fly ash deposit site using surface electrical resistivity methods. Ground Water, **18**, (5), 472-478.

Roka, A., and Tsokas, G. (1987). Field investigation of a Macedonian tumulus by resistivity soundings. Geoexploration, **24**, 99-108.

Rothman, D.H. (1985). Nonlinear inversion, statistical mechanics and residual statics estimation. Geophysics, **50**, 2784-2796.

Roy, A., and Apparao, A. (1970). Depth of investigation in direct current methods. Geophysics, **36**, 943-959.

Roy, A. (1971). Depth of investigation in Wenner, three-electrode and dipole-dipole DC eesistivity methods. Geophysical Prospecting, **20** ,29-340.

Sasaki, Y. (1982). Automatic inversion of induced polarization data over 2-D struc-

tures. Memoirs of the Faculty of Engineering, Kyushu University, **42**, 59-74.

Sasaki, Y. (1989). 2-D joint inversion of magnetotelluric and dipole-dipole resistivity data. Geophysics, **54**, 254-262.

Sasaki, Y (1992). Resolution of resistivity tomography inferred from numerical simulation. Geophysical prospecting, **40**, 453-464.

Sen, M.K. Bhattacarya, B.B., and Stoffa, P.L. (1993) Nonlinear inversion of resistivity sounding data. Geophysics, **58**, 496-507.

Shima, H. (1990). Two-dimensional automatic resistivity inversion technique using alpha centres. Geophysics, **55**, 682-694.

Shima, H. (1992). 2-D and 3-D resistivity image reconstruction using crosshole data. Geophysics, **57**, 1270-1281.

Skilling, J. (1988). The axioms of maximum entropy: in Maximum Entropy and Bayesian Methods in Science and Engineering, Vol. 1 Foundations, Erickson, G.J. and Smith C.R. (eds.), Kluwer.

Skilling, J. (1989). Classic maximum entropy: in Maximum Entropy and Bayesian Methods ,Skilling, J. (ed.), Kluwer.

Smith, D. (1974). Variational methods in optimization. Pentice-Hall.

Smith, N., and Vozoff, K. (1984). Two-dimensional DC resistivity inversion for dipole-dipole data. IEEE Trans. Geosc., **22**, (1), 21-28

Smith, D. (1986). Application of the pole-dipole resistivity technique to the detection of solution cavities beneath the highways. Geophysics, **51**, 833-837.

Stretenovic, B., and Marcetic, D. (1992). Determination of the internal geometry of a land-slide using electrical tomography. Abstracts of the 54th Meeting of the E.A.E.G., Paris, France, 1-5 June.

Sumner, J. (1976). Principles of induced polarization for Geophysical Exploration.

Elsevier, Amsterdam.

Szymanski, J., and Dittmer, J. (1992) Inversion of archaeological magnetic survey data using the maximum entropy method: Preliminary Results. Monograph on Geophysical Data Inversion in Archaeological Site Investigation, Vieweg Verlag (in press).

Szymanski J., Cambell T., Dittmer J, Giannopoulos A., Tsourlos P., Coppack P., Emerick K. and Wilson K. (1992). Non-destructive site diagnosis at medieval abbey sites in the UK. Proceeding of MEDIEVAL EUROPE 1992, York, U.K., 21-24 September 1992, 201-206.

Szymanski, J.E., Tsourlos, P., and Dittmer, J. (1994) Maximum entropy reconstruction of resistivity data-sets using a finite element forward model. Expanded abstracts of the 56th meeting and technical exhibition of the EAEG: Vienna, Austria, 6-10 June, 1994, P128.

Taflove, A. (1995). Computational electrodynamics: the FDTD method. Artech House.

Tagg, G. (1964). Earth Resistances. George Newnes Ltd.

Tarantola, A., and Valette, B. (1982). Generalized nonlinear inverse problems solved using the least squares criterion. Rev. Geophys. and Space Phys., **20**, (2), 219-232.

Tarantola, A. (1987). Inverse problem theory. Elsevier, Amsterdam.

Tarassenko, L., and Rolfe, P. (1984) Imaging spatial distributions of resistivity, an alternative approach. Elec. Lett., **20**, 574-575.

Telford, W., Geldart, L., Sheriff, R., and Keys, D. (1991). Applied geophysics. Cambridge University Press.

Thanassoulas, and Tsokas, G. (1987). Geophysical investigation in the geothermal field in the delta of Nestos river (N. Greece). Geothermics, 16, 1, 17-26.

Tikhonov, A.N. (1963). Solution of incorrectly formulated problems and the regu-

larization method. Soviet Mathematics, **4**, 1035-1038.

Tong, L., and Yang, C. (1990). Incorporation of topography into 2-D resistivity inversion. Geophysics, **55**, 354-361.

Tripp, A., Hohmann, G., and Swift, C. (1984). Two-dimensional resistivity inversion. Geophysics, **49**, 1708-1717.

Tsokas, G., and Papazachos, C. (1992). Two-dimensional inversion filters in magnetic prospecting: Application to the exploration of buried antiquities. Geophysics, **57**, 1004-1013.

Tsokas G., Giannopoulos A., Tsourlos P., Vargemezis J., Tealby J., Sarris A., Papazachos C., and Savopoulou T. (1994). A large scale geophysical survey in the archaeological site of Europos (N.Greece). Journal of Applied Geophysics, **32**, 85-98.

Tsourlos P. (1993). Modelling and inversion schemes for vertical profiling resistivity data. Internal report, Department of Electronics, University of York.

Tsourlos P., Szymanski J., Dittmer J., and Tsokas G. (1993). The use of back-projection for fast inversion of 2-D resistivity data. Proceedings of the 2nd congress of the Greek Geophysical Union, Florina, Greece, 5-7 May, 1993, vol. 1, 71-81.

Tsourlos, P., Dittmer, J., and Szymanski, J. (1995). A study of non-linear techniques for the 2-D inversion of earth resistivity data. Expanded abstracts of the 57th meeting and technical exhibition of the EAEG: Glasgow, Scotland, 29 May- 2 June, 1995, P031.

van Laarhoven, P.J.M., and Aarts, E.H.L. (1987). Simulated annealing: theory and applications. D.Reidel Publishing Company.

Van, G.P. Park, S.K., and Hamilton, P. (1992). Use of resistivity monitoring systems to detect leaks from storage ponds. Proceedings of SAGEEP 92, Chicago, 629-647.

Van Dam, J. (1976). Possibilities and limitations of the resistivity method for geo-

electrical prospecting in the solution of geohydrological problems. Geoexploration, **14**, 179-193.

Vandenberghe, J. (1982). Geoelectric investigations of a fault system in quartenary deposits. Geophysical Prospecting, **42**, 977-991.

Vaughan, L. (1986). Ground penetrating radar surveys in archaeological investigations. Geophysics, **51**, 595-604.

Wang, M., Dickin, F.J., and Williams, R.A. (1994). Electrical resistance tomography on metal walled vessels. Proceedings of ECAPT'94 Process Tomography, Oporto, 163-170.

Ward S., and Hohmann G., (1988). Electromagnetic theory for geophysical applications: in Electromagnetic Methods in Applied Geophysics, Nabighian M.N. (ed.) vol 1, Ch. 4. SEG, Tulsa, 131-312.

Ward, S (1989) Resistivity and induced polarization methods: in Investigations in Geophysics no 5, Geotechnical and Environmental Geophysics vol I, ed. S. Ward, SEG, Tulsa, 147-189.

Weymouth, J. (1986). Geophysical methods in archaeological site surveying. Advances in Archaeological Method and Theory, **9**, 311-395.

Wright, P.M., Ward, S.H., Ross, H.P., and West R.C. (1985). State-of-the-art geophysical exploration for geothermal resources. Geohysics, **50**, 2666-2699.

Xu, B., and Noel, M. (1993). On the completeness of data sets with multielectrode systems. Geophysical Prospecting, **41**, 791-801.

Xu, B. (1993). Development of electrical resistivity imaging methods for geological and archaeological prospecting. Ph.D. Thesis, University of Durham.

Yorkey, T. (1986). Comparing rReconstruction methods for electrical impedance tomography. Ph.D. Thesis. University of Wisconsin - Madison.

Yorkey, T., Webster, J., and Tompkins, W. (1987a). Comparing reconstruction

methods for electrical impedance tomography. IEEE Trans. Biom. Eng., **34**, 843-852.

Yorkey, T., Webster, J., and Tompkins, W. (1987b). An improved perturbation technique for electrical impedance tomography with some criticisms. IEEE Trans. Biom. Eng., **34**, 898-901.

Zhang J., Mackie, R.L., and Madden T.R. (1995). 3-D resistivity forward modelling and inversion using conjugate gradients. Geophysics, **60**, 1313-1325.

Zhao, J.S., Rijo, L., Ward, S.H. (1986). Effects of geological noise on cross-borehole electrical surveys. Geophysics, **51**, 1978-1991.

Zhody, A. (1989). A new method for the interpretation of Schlumberger and Wenner sounding curves. Geophysics, **54**, 245-253.

Zienkiewicz, O.C., and Taylor, R.L. (1989). The finite element method. 4th ed., Vol. 1, Basic formulation and linear problems. McGraw-Hill.

Appendix A: 2DINVS-User's manual

A1. ABOUT THE PROGRAM

2DINVS is a program for 2-D inversion of surface earth resistivity data which operates under DOS. The program performs smoothness constrained (Occam's) inversion and is based on a finite element forward modelling scheme. The Jacobian matrix is calculated using the Adjoint equation technique. The main features of the program are:

- It can reconstruct data obtained using several known arrays. It can cope with a large amount of data since it uses a DOS extender.
- There is an option for using Quasi-Newton update of the Jacobian matrix in order to speed up the inversion.
- The parameters of the problem can either be decided automatically or by the user (till a certain degree).
- Some options for adjusting the model smoothness are included.

A2. HOW TO INSTAL AND RUN IT

Copy the contents of the disk to a directory. Within the disk there are:

a. the [2DINVS.EXE] which is the executable inversion program

b. the [GO32.EXE] file which is the DOS-extender [1]. Note that unless [GO32.EXE] is in the same directory with the [2DINVS.EXE] or in the DOS-Path the program will not run.

c. Sample inversion input files and some test data files.

In order to run the program you type: 2DINVS <*filename*> where filename is the name of a 2DINVS inversion input file.

Example: > `2DINVS TST.IN`

or >`2DINVS HILL.DAT`

The format of the inversion input file is explained in Section A3.

A3. THE INVERSION INPUT FILE

The inversion input file contains all the information that the inversion program needs in order to run. The input file can have any name however it should have a specific format. It consists of several lines. If the line starts with the symbol $ then this line is a header which is actually ignored by the program and is just for the operator's guidance, however this line should always be in the file. All other lines (without the $ symbol) should contain the required values or filenames. Note that the sequence of the lines must never change.

The first line of the input file is a general header:

INVERSION_INPUT_FILE_FOR_"2DINVS"

It is followed by the line:

$----------

this type of line is an indication of a different **part** of the input file. The input file has three parts:

PART 1: Includes information for the dynamic memory allocation.

PART 2: Includes information about the type of the resistivity data to be inverted

[1]GO32.EXE is a public domain DOS extender (Free Software Foundation Inc.).

as well as info about the output.

PART 3: Includes information used by the inversion algorithm.

An example of an input file is shown in Figure A1.

The options that have to be inserted are described below:

PART 1

$Max_probes

integer *max_probes*

max_probes= The maximum number of probes (electrodes) that the program can handle. This value is used exclusively for memory allocation purposes and therefore any value equal or larger than the actual number of electrodes of the inverted data-set is valid

Example:

```
$Max_probes
20
```

This is a proper value for a data-set of 1-20 electrodes.

$Max_n

integer *max_n*

max_n= The maximum probe separation (expressed in integer multiples of the unit electrode spacing) that the program can handle. This value is the well-known n-separation for the Wenner, Dipole-dipole pseudosection and similarly can be defined for other arrays. This value is used exclusively for memory allocation purposes and therefore any value equal or larger than the actual number of the n-separation of the inverted data-set is valid.

```
$INVERSION_INPUT_FILE_FOR_"2DINVS"
$-----------------------------------------------------------------
$MAX_PROBES
20
$MAX_N
8
$MAX_MEASUREMENTS
105
$-----------------------------------------------------------------
$DATA_TYPE_(Apres=1,Resistance=2,Apres+noise=11,Resistance+noise=12)
11
$ARRAY_TYPE_(Wen=1,Dip-Dip=2,Pole-Dip=3,Pole-Pole=4)
2
$DATA_INPUT_FILENAME
test.dat
$INFO_OUTPUT_FILENAME
test.inf
$-----------------------------------------------------------------
$MAXIMUM_NUMBER_OF_ITERATIONS
2
$INVERSION_TYPE_(Quasi-Newton=1,Newton=2)
1
$LAGRANIAN_MULTIPLIER
0.5
$SMOOTHNESS_TYPE_(normal=1,>bottom=2)
1
$DEFINE_PARAMETERS_(auto=0,user=1)
1
$PARAMETER_USER_FILENAME
test.prm
$END
```

Figure A1: An example input file

Example:

$Max_n

7

This is a proper value for a data-set of 1-7 n-separation.

$Max_measurements

integer *max_meas*

max_meas= The maximum number of measurements that the program can handle. This value is used exclusively for memory allocation purposes and therefore any value equal or larger than the actual number of the measurements of the inverted data-set is valid.

Example:

$Max_measurements

55

This is a proper value for a data-set of 1-55 measurements.

PART 2

$DATA_TYPE

integer *data_type*

data_type=1, or 2, or 11, or 12

Is a number which indicates the type of the measured data inserted into the program. [The full format that the data should have is given in Section A4]. The numbers that can be given are:

1 if the data are **apparent resistivities** (in Ohm-m) and **no** data errors are given.

2 if the data are **resistances** (in Ohms) and **no** data errors are given.

11 if the data are **apparent resistivities** (in Ohm-m) and data errors are also given (in Ohm-m).

12 if the data are **resistances** (in Ohms) and data errors are also given (in Ohms).

Example:

```
$DATA_TYPE
1
```

In this case the given data are apparent resistivities in Ohm-m

$ARRAY_TYPE

integer *data_type*

data_type=1, or 2, or 3, or 4

Is a number which indicates the type of the array used to obtain the measured data. The numbers that can be given are:

1 for Wenner array

2 for Dipole-dipole array

3 for Pole-dipole array

4 for Pole-pole array

Example:

```
$ARRAY_TYPE
4
```

In this case the given data are obtained using the Pole-pole array.

$DATA_INPUT_FILENAME

string *in_name*

in_name= The DOS name of the file which contains the data to be inverted. [The full format that the data should have is given in Section A4].

Example:

```
$DATA_INPUT_FILENAME
test.dat
```

$INFO_OUTPUT_FILENAME

string *out_name*

out_name= The DOS name of the file in which the inversion information and results will be stored. [The full format of the output info file is given in Section A5].

Example:

```
$INFO_OUTPUT_FILENAME
out.dat
```

PART 3

$MAXIMUM_NUMBER_OF_ITERATIONS

integer *max_itr*

max_itr= The maximum number of iterations that the program is allowed to perform. Usually the program will converge in few (less than 6) iterations. Since the program has other stopping criteria if you do not want to use number of iterations as a stopping criterion put a large number (e.g. 20) as a value.

Example:

```
$MAXIMUM_NUMBER_OF_ITERATIONS
10
```

The program will not go further than 10 iterations.

$INVERSION_TYPE

integer *inv_type*

inv_type=1, or 2

Is a number which indicates the type of the inversion the program will perform. The numbers that can be given are:

1	for smoothness constrained Quasi-Newton inversion (quasi-Newton update of the Jacobian matrix).
2	for smoothness constrained Newton-type inversion (calculation of the Jacobian at every iteration.

Example:

```
$INVERSION_TYPE
1
```

In this case Quasi-Newton inversion (quasi-Newton update of the Jacobian matrix) will be performed.

$LAGRANIAN_MULTIPLIER

float *lagr_mlt*

lag_mlt= A float value of the Lagranian multiplier. A typical value is 0.5 but this can change according to your data set.

Example:

```
$LAGRANIAN_MULTIPLIER
0.5
```

$SMOOTHNESS_TYPE

integer *smooth_type*

smooth_type=1, or 2,
Is a number which indicates the parameter smoothness scheme to be followed. The numbers that can be given are:

1	for a uniform (equal in all directions) smoothness scheme.
2	for increasing smoothness with depth.

Example:

```
$SMOOTHNESS_TYPE
2
```

This will create increased smoothness with depth.

$DEFINE_PARAMETERS

integer *def_param*

def_param=0, or 1

Is a number which indicates the way the parametrization will happen (see Section 6.5). The numbers that can be given are:

0	the program will do the parametrization automatically.
1	the program will do the parametrization based on user defined parameters - a filename has to be also defined using the option below.

Example:

```
$DEFINE_PARAMETERS
1
```

$PARAMETER_USER_FILENAME

string *param_name*

param_name= The DOS name of the file which contains the parametrization information. Needed only if the DEFINE_PARAMETERS option is 1. The general format of the file is given in section A6.

Example:

`$PARAMETER_USER_FILENAME`

`test.prm` or `prm.dat`

A4 FORMAT OF THE DATA INPUT FILE

The data input file can have any Dos name however it should have a specific format. The data input file is a **series of lines**. Each line bears information for **one** measurement of the data-set. The format of the line depends on the choice of the `$DATA_TYPE` option:

If *data_type=1* (apparent resistivities without errors) a line of the data file should have the following format:

A_elec_pos B_elec_pos M_elec_pos N_elec_pos ap_res

If *data_type=2* (resisances without errors) a line of the data file should have the following format:

A_elec_pos B_elec_pos M_elec_pos N_elec_pos restanc

If *data_type=11* (apparent resistivities with errors) a line of the data file should have the following format:

A_elec_pos B_elec_pos M_elec_pos N_elec_pos ap_res apres_error

If *data_type=11* (resistances with errors) a line of the data file should have the following format:

A_elec_pos B_elec_pos M_elec_pos N_elec_pos resstanc restanc_error

where:

A_elec_pos is the location (x coordinate) in metres of the current electrode A (I^+) for a measurement.

B_elec_pos is the location (x coordinate) in metres of the current electrode B (I^-) for a measurement.

M_elec_pos is the location (x coordinate) in metres of the potential electrode M (V^+) for a measurement.

N_elec_pos is the location (x coordinate) in metres of the potential electrode N (V^+) for a measurement.

ap_res the measured apparent resistivity in Ohm-m.

ap_res_error the standard deviation of the measurement's error in Ohm-m.

restanc the measured resistance in Ohms.

restanc_error the standard deviation of the measurement's error in Ohms.

Example:

Suppose that a full Wenner data set is obtained over a homogeneous half space of 100 Ohm-m. The 7 probes used are positioned at 10, 20, 30, ...70m. The input data file will be (depending on the *data_type* choice)

data_type=1:

```
10.0 40.0 20.0 30.0 100.0
20.0 50.0 30.0 40.0 100.0
30.0 60.0 40.0 50.0 100.0
40.0 70.0 50.0 60.0 100.0
10.0 70.0 30.0 50.0 100.0
```

data_type=2:

```
10.0 40.0 20.0 30.0 1.591
20.0 50.0 30.0 40.0 1.591
30.0 60.0 40.0 50.0 1.591
40.0 70.0 50.0 60.0 1.591
10.0 70.0 30.0 50.0 1.591
```

data_type=11 (5% error):

```
10.0 40.0 20.0 30.0 100.0 5.0
20.0 50.0 30.0 40.0 100.0 5.0
30.0 60.0 40.0 50.0 100.0 5.0
40.0 70.0 50.0 60.0 100.0 5.0
```

```
10.0 70.0 30.0 50.0 100.0 5.0
```

data_type=22 (5% error):

```
10.0 40.0 20.0 30.0 1.591 0.08
20.0 50.0 30.0 40.0 1.591 0.08
30.0 60.0 40.0 50.0 1.591 0.08
40.0 70.0 50.0 60.0 1.591 0.08
10.0 70.0 30.0 50.0 1.591 0.08
```

The following have to be taken in to account when creating the data input file:

1. All columns of the file should be separated by a space (**no commas**)[2].

2. The measurements can be inserted in any sequence.

3. You should **never** use **negative** coordinates for the probes [3].

4. In case you want to enter arrays that are using electrodes at "infinity" (pole-dipole and pole-pole) the format of the file should remain the same - the program will simply ignore the coordinates of electrodes at "infinity". Example:

The correct pole-pole measurement entry with probes A,M at 15, 25m respectively (for the half-space of 100 Ohm-m) should be:

```
15.0 0.0 25.0 0.0 100.0
```
or
```
15.0 500.0 25.0 0.0 100.0
```
or
```
15.0 100.0 25.0 220.0 100.0
```

the following is incorrect:

```
15.0 25.0 100.0
```

[2]There is no need to insert a number as a float if it has no decimal digits. For example it is equally correct to enter an electrode positioned of 10m as 10.0 or 10

[3]Sometimes negative coordinates are used since it is common to use the centre of the measured area as the beginning (0m) of the coordinate system. In such a case you should render all electrode coordinates in positive- by adding the appropriate positive number- before you invert the data. Subsequently you can subtract this number from the inversion results in order to return to the original coordinate system

A5 FORMAT OF THE OUTPUT FILE

The output file can have any name. It contains all information about the inversion results. The File has 3 parts:

PART 1

The first part is a general header which keeps a record of the selected options.

PART 2

The second part gives information about the progress of the inversion. It presents the results at each iteration in a pseudosection-like form [4]. It keeps track of the lagranian multiplier and the % normalized RMS for each iteration.

PART 3

The third part gives the final inversion results:

——— ERRORS ———————

3 lines of the variation of the RMS error through the iterations.

———PSEUDOSECTION OF REAL-MODEL DATA———

a <*x y z1 z2*> file containing the pseudosections of the real and model data for the last iteration: *x*=the x-coordinate of the pseudosection point, *y*=the y-coordinate of the pseudosection point, *z1*= the measured (observed) data, *z2*=the calculated data

———— XYZ FILE ———————

a < *x y z* > file containing the resistivities of the parameters: *x*=the x-coordinate of the centre of the parameter, *y*=the y-coordinate of the centre of the parameter, *z*=the resistivity of the parameter.

A6. PARAMETER_USER_FILENAME

The parameter user file can have any DOS name however it should have a specific format:

[4]In order to facilitate presentation the results are presented as integers so if you see 0 resistivity values that means that the value is below 1. For large numbers of parameters this facility becomes effectively redundant since there is not enough space for writing everything in order.

line 1:	*int x_spacing int numl*
line 2:	*float thick_layer_1*
line 3:	*float thick_layer_2*
.	.
.	.
.	.
line numl+1	: *float thick_layer_numl*

where:

x_spacing is the width (x-direction) of the parameters in **inter-electrode spacing units**. It can **only** have two values: **1** or **2**.

numl is the number of layers the parametrized space will have.

thick_layer_1,2...numl are the thicknesses of the 1st,2nd..numl layers respectively in metres.

Example:

```
1 2
10
20
```

A7. FINAL REMARKS

- The program cannot cope with irregular spaced electrodes.
- The program will not do automatically topographic corrections. You have to remove topographic effects using an other program before you invert the data.
- This program cannot cope with tomographic-type data.
- Although the program have some basic controls in general you have to be

careful that your selected options are consistent with the data. For example if you are inverting pole-pole data and the ARRAY_TYPE option is not set to 4 then the program will either crush or will produce nonsense.

Therefore if you are facing problems: a) check the validity of the chosen options in the inversion input file. b) check that your input data file is correct.

If you are still getting meaningless or unstable results try to increase the Lagranian multiplier- maybe its value is too low.