THEORY AND APPLICATION OF JOINT INTERPRETATION OF MULTIMETHOD GEOPHYSICAL DATA

ELENA KOZLOVSKAYA

Department of Geophysics,
University of Oulu

OULU 2001

ELENA KOZLOVSKAYA

THEORY AND APPLICATION OF JOINT INTERPRETATION OF MULTIMETHOD GEOPHYSICAL DATA

Academic Dissertation to be presented with the assent of the Faculty of Science, University of Oulu, for public discussion in Kajaaninsali (Auditorium L6), Linnanmaa, on May 25th, 2001, at 12 noon.

OULUN YLIOPISTO, OULU 2001

Manuscript received 4 April 2001
Manuscript accepted 12 April 2001

Communicated by
Professor Hermann Zeyen
Professor Karsten Bahr

ISBN 951-42-5960-2 (URL: http://herkules.oulu.fi/isbn9514259602/)

ALSO AVAILABLE IN PRINTED FORMAT
ISBN 951-42-5959-9
ISSN 0355-3191 (URL: http://herkules.oulu.fi/issn03553191/)

OULU UNIVERSITY PRESS
OULU 2001

Kozlovskaya, Elena, Theory and application of joint interpretation of multimethod geophysical data

Department of Geophysics, University of Oulu, P.O.Box 3000, FIN-90014 University of Oulu, Finland

2001

Oulu, Finland

(Manuscript received 4 April 2001)

Abstract

This work is devoted to the theory of joint interpretation of multimethod geophysical data and its application to the solution of real geophysical inverse problems. The targets of such joint interpretation can be geological bodies with an established dependence between various physical properties that cause anomalies in several geophysical fields (geophysical multiresponse). The establishing of the relationship connecting the various physical properties is therefore a necessary first step in any joint interpretation procedure. Bodies for which the established relationship between physical properties is violated (single-response bodies) can be targets of separate interpretations. The probabilistic (Bayesian) approach provides the necessary formalism for addressing the problem of the joint inversion of multimethod geophysical data, which can be non-linear and have a non-unique solution. Analysis of the lower limit of resolution of the non-linear problem of joint inversion using the definition of e-entropy demonstrates that joint inversion of multimethod geophysical data can reduce non-uniqueness in real geophysical inverse problems. The question can be formulated as a multiobjective optimisation problem (MOP), enabling the numerical methods of this theory to be employed for the purpose of geophysical data inversion and for developing computer algorithms capable of solving highly non-linear problems. An example of such a problem is magnetotelluric impedance tensor inversion with the aim of obtaining a 3-D resistivity distribution. An additional area of application for multiobjective optimisation can be the combination of various types of uncertain information (probabilistic and non-probabilistic) in a common inversion scheme applicable to geophysical inverse problems. It is demonstrated how the relationship between seismic velocity and density can be used to construct an algorithm for the joint interpretation of gravity and seismic wide-angle reflection and refraction data. The relationship between the elastic and electrical properties of rocks, which is a necessary condition for the joint inversion of data obtained by seismic and electromagnetic methods, can be established for solid- liquid rock mixtures using theoretical modelling of the elastic and electrical properties of rocks with a fractal microstructure and from analyses of petrophysical data and borehole log data.

Keywords: joint inversion, multiobjective optimisation, non-probabilistic uncertainty, fractal rock models

Acknowledgements

I am very grateful to many people who have helped me to complete this work, especially to those whom I regard as having been my main teachers in geophysics.

I am particularly indebted to my first scientific adviser, Prof. German Karatayev, who invited me to study the problem of the joint interpretation of multimethod geophysical data at the Institute of Geological Sciences in the Academy of Sciences of Belarus. The method of joint interpretation of seismic and gravity data on the basis of the density-velocity relationship was developed under his supervision.

I would like to thank very much my teacher and co-author, the seismologist Jukka Yliniemi of Sodankylä Geophysical Observatory, University of Oulu, who raised the idea of applying the method of joint interpretation of gravity and seismic data to the EUROBRIDGE and SVEKA wide-angle reflection and refraction profiles and developed the high quality seismic models for joint interpretation. His essential recommendations concerning different aspects of seismic data interpretation were of great importance for my work.

I am particularly grateful to Professor Sven-Erik Hjelt who invited me to apply and develop my ideas for joint inversion of seismic and electromagnetic data within the NOIGEM and SVEKALAPKO projects. He introduced me to a variety of interesting problems concerning electromagnetic data inversion and proposed the use of fractal models for investigating the relationship between the elastic and electrical properties of lithospheric rocks. He helped me with critical comments and advice at every stage in my research.

I particularly wish to thank all my colleagues at the Department of Geophysics, University of Oulu, for the pleasant working atmosphere at the department and for their critical comments on my work. The department's computer specialist, Kari Komminaho, helped me with all my problems concerning the proper functioning of the hardware and concerning computer programming. Riitta Hurskainen, software specialist at the Oulu Unit of Sodankylä Geophysical Observatory was always ready to assist in solving technical problems.

I am very grateful to Professor Germann Zeyen and to Professor Karsten Bahr for their careful review of this thesis. I would also like to thank Mr. Malcolm Hicks, M.A., of his efforts to improve the language and style of the manuscript.

Finally, I would like to say many thanks to my family for their patience, and especially to my husband, Dr. A. Kozlovsky, for his qualified assistance with the theory of electromagnetic wave propagation.

The research was supported financially by the Academy of Finland.

Oulu, March, 2001 Elena Kozlovskaya

List of original papers

This thesis is based on the following four original papers, which are referred to in the text by their Roman numerals:

I Kozlovskaya, E. (2000) An algorithm of geophysical data inversion based on non-probabilistic presentation of a-priori information and definition of Pareto-optimality. Inverse Problems, 16: 839-861.

II Kozlovskaya, E., Karatayev, G., Yliniemi, J. (2001) Lithosphere structure along the northern part of EUROBRIDGE in Lithuania; results from integrative interpretation of DSS and gravity data. Tectonophysics (in press).

III Kozlovskaya, E., Yliniemi, J. (1999) Deep structure of the Earth's crust along the SVEKA profile and its extension to the north-east. Geophysica, 35, 1-2: 111-123.

IV Kozlovskaya, E., Hjelt, S.-E. (2000) Modelling of elastic and electrical properties of solid-liquid rock system with fractal microstructure. Physics and Chemistry of the Earth (A), 25, 2: 195-200.

The thesis also contains unpublished results.

Contents

Introduction

At the present moment it is clear to everyone dealing with the interpretation of geophysical data that the perfect method for probing the deep structure of the Earth does not exist. Every method has its own strengths and weaknesses, and only a combination of various methods can produce a realistic model of the Earth's interior. Unfortunately, a very common practice in multidisciplinary geophysical projects is that specialists in each geophysical method form a research group of their own at the beginning of the project and interpret the experimental results in isolation from the other specialists. At the final stage all the specialists get together to compile a combined model, and are very often faced with the problem that the models produced by the different groups without taking into consideration data derived from other methods cannot be combined as they contradict each other.

One possible way to avoid this problem is to make the combined model not at the end of interpretation, when a great amount of work has already been completed by specialists in all the groups, but from the very beginning, i.e. to perform *joint interpretation of multimethod geophysical data*. The problem is far from a simple one, however, and it cannot be solved if one just puts all the experimental data into one common data set and inverts it using a powerful computer.

I began to study the problem of joint interpretation of geophysical data sets systematically in 1989 during my postgraduate work at the Institute of Geological Sciences of the Academy of Sciences of Belarus in Minsk, under supervision of Prof. G. Karatayev. At that time the problem of joint interpretation was a topic of investigation of several research groups in the former Soviet Union. Two of the approaches developed then were selected as a background to real algorithms for the joint interpretation of several geophysical data sets.

The first was a statistical approach (Bayesian approach) developed by Prof. F.M. Goltsman and T.B. Kalinina at St. Petersburg University since the 1960's. This enables joint inversion of several geophysical experimental data sets and assumes that the unknown geological medium can be described by a parameter vector common to all the data sets that allows a generalised objective function to be constructed in a natural and convenient way as a joint *a-posteriori* probability density function of the observed data. The book by Goltsman (1982) contains a complete description of the approach, and thus

reference is made to that book in the present work, although a number of earlier papers published in Russian editions also exist.

The second approach is that proposed by Prof. G.I. Karatayev, and described by him as a "correlation theory of integrated interpretation of geophysical fields". It is based on the supposition that there exists and can be found in principle an equation connecting various physical parameters of geological media, e.g. density and seismic velocity. The main point in this approach is use of the relationship between various physical parameters of geological media for the joint interpretation of multimethod geophysical data. This unknown relationship is approximated by a function constructed from *a-priori* information. It should be noted that the idea of the method was proposed by Prof. G.I. Karatayev in 1980s (Karatayev and Pashkevich, 1986), but could not be implemented at that time due to the lack of sufficiently powerful computers.

In my Candidate in Science thesis written in Minsk I considered these two approaches, formulated a theoretical background to the joint inversion of multimethod geophysical data on the base of multiobjective optimisation theory and developed corresponding computer programs for the joint inversion of gravity and magnetic data. I also developed computer algorithms for solving the problem of joint interpretation of seismic and gravity data using the relationship between density and seismic velocity.

There was a difficult time in the history of many of the republics of the former Soviet Union when interesting research topics became frozen for many years, but eventually the INTAS programme of the European Science Foundation helped some researchers to continue their studies. The multidisciplinary EUROPROBE/EUROBRIDGE/INTAS project carried out by an international research team in Belarus gave me the possibility to continue my investigations and apply my previously developed techniques of joint interpretation of seismic and gravity data to the good quality experimental data obtained in the EUROBRIDGE wide-angle reflection and refraction profiles. Subsequent work at the Oulu Unit of Sodankylä Geophysical Observatory, in collaboration with the seismologist Jukka Yliniemi, and then at the Department of Geophysics of Oulu University under the supervision of Prof. Sven-Erik Hjelt on the problem of joint inversion of seismic and electromagnetic data marked a continuation of my previous investigations in the area of joint interpretation.

The present work consists of four chapters. Chapter 1 is based on my former Candidate in Science thesis and articles devoted to certain theoretical aspects of the joint interpretation of multimethod geophysical data. Several of these articles were written in Russian and published by the Academy of Sciences of Belarus. Selected conclusions from those papers are incorporated into the present work. Section 1.2 contains an analysis of non-uniqueness and resolution of the problem of non-linear joint inversion of multimethod geophysical data. It is demonstrated in Section 1.3 that the joint inversion problem can be regarded as a problem of multiobjective optimisation. Multiobjective formulation allows application of the numerical methods of multiobjective optimisation for the purpose of geophysical data inversion and the development of computer algorithms capable to solve strongly non-linear problems. Some of the basics of the theory of multiobjective optimisation are also included in this part to help the reader to understand the problem.

Section 1.4 is devoted to the problem of presenting *a-priori* information in geophysical inverse problems. As the non-probabilistic theories of uncertainty are not

very popular among geophysicists, I have included a discussion of various types of uncertainty in geophysical inverse problems in this section. This question is also considered in Paper I, where it is demonstrated that the use of measures of uncertainty other than probabilistic ones can be an efficient tool for presenting *a-priori* information in geophysical inverse problems. A review of the main non-probabilistic measures of uncertainty is given in Appendix 1 to help the reader to understand the problem. Appendix 2 contains the main aspects of fuzzy set theory.

Chapter 2 is devoted to the joint interpretation of seismic and gravity data using an equation connecting density with the velocity of elastic waves, and is based on Papers II and III, in which the method itself and some results of its application to the joint interpretation of seismic and gravity data along the EUROBRIDGE'95 (Lithuania) and SVEKA (Finland) wide-angle reflection and refraction profiles are discussed. The seismic models described in these papers were developed by J. Yliniemi.

Chapter 3 is devoted to the relationship between elastic and electrical rock properties. The establishing of this relationship is the main condition for the joint inversion of seismic and EM data. The results of the petrophysical investigations contained in the KTB deep drilling programme and theoretical calculations of rock elastic and electrical properties were used for this purpose. Section 3.1 contains an analysis of previous theoretical investigations into common factors that affect both elastic and electrical properties. The theoretical rock model that allows elastic and electrical properties of solid-liquid rock systems with a fractal microstructure to be calculated simultaneously is discussed in Section 3.2 and Paper IV. Selected results of calculations of elastic properties of crustal rock by means of the fractal rock model are described in Section 3.2. Section 3.3 contains an analysis of the borehole evidence for a relationship between velocity and electrical resistivity and of the multiscale nature of the factors affecting these two properties in the upper crust. This leads to the conclusion that a multiscale presentation of geological media can be used to separate the factors affecting various physical properties of rocks and can become an additional tool for use in the joint inversion of multimethod geophysical data.

The problem of magnetotelluric impedance tensor inversion is dealt with in Chapter 4, where it is demonstrated that multicritera optimisation, i.e. the ideal point method, can be used to develop robust algorithms for the inversion of magnetotelluric impedance tensor data.

The main results are summarised in the Conclusions section.

1 Theory of joint interpretation of multimethod geophysical data

1.1 Relationship between physical properties as a major condition for the joint interpretation of multimethod geophysical data

The physical background to the joint interpretation of multimethod geophysical data is the empirical fact that there exists a large class of geological objects that cause anomalies in several geophysical fields. Ore deposits are often sources of both anomalous gravity and magnetic fields, conductive zones in the crust can coincide with areas of high seismic reflectivity, and a close correlation exists between density and compressional wave velocity in crustal rocks etc.

Consider a geological body T limited by a surface S and having physical parameters (density, magnetisation, P- and S-wave seismic velocity etc.) denoted as $\beta_1(T)$, $\beta_2(T)$,…, $\beta_k(T)$. If a physical property of the body T, namely, $\beta_i(T)$ differs from that of the surrounding medium, it should cause an anomaly in the corresponding geophysical fields that can be expressed as

$$U_i(\mathbf{r}) = F_i\left[\beta_i(\mathbf{t}), T\right], \mathbf{t} \in T, \ \mathbf{r} \in Q, i = 1,2,...,k, \tag{1.1}$$

where Q is the observation area, the value $\beta_i(\mathbf{t})$ implies the ith parameter of the body T and $F_i(\cdot)$ is an operator (non-linear in the most common case) connecting the ith geophysical field to the corresponding physical parameter. Bodies with anomalous physical properties that cause observable anomalies in geophysical fields are the targets of geophysical investigations.

Assume that the formation of the body T was controlled by a number of physical and geological processes that influenced various physical properties simultaneously. As a result, the physical parameters of the body T assumed a certain relationship one to another. This relationship is probabilistic in nature and can be formally described by a function, e.g. a probability density function, regression equation etc.

$$W[\beta_1(\mathbf{t}), \beta_2(\mathbf{t}),..., \beta_k(\mathbf{t})] = 0, \ \ \mathbf{t} \in T \, . \tag{1.2}$$

The geological body T with physical properties $\beta_i(\mathbf{t}), i = 1,2,..., k$, connected by equation (1.2) may be referred to as *a multiresponse geophysical body* (Karatayev and Kozlovskaya, 1997).

If the physical parameters of the body T are connected, i.e. they satisfy equation (1.2), this will cause anomalies in the corresponding geophysical fields which can be expressed as a system of integral and differential equations that are commonly non-linear:

$$\begin{cases} U_1(\mathbf{r}) = F_1[\beta_1(\mathbf{t}), T], \\ \qquad ... \qquad\qquad\qquad \mathbf{t} \in T, \ \mathbf{r} \in Q. \\ U_k(\mathbf{r}) = F_k[\beta_k(\mathbf{t}), T], \end{cases} \tag{1.3}$$

The system of equations (1.3) can be solved in principle for the unknown physical parameters $\beta_i(\mathbf{t}), i = 1,2,..., k$ and the geometry of the body T, assuming that the fields in the observation area Q denoted as $U_i^{obs}(\mathbf{r}), i = 1,2,..., k$ are known from measurements. This procedure can be called the *joint interpretation of multimethod geophysical data.* It is necessary to stress that if a physical parameter of the body T does not differ from that of the surrounding media, the corresponding geophysical field in the observation area will not differ from the background "normal" value, i.e. it will not contain information about the body T. This situation can occur if some of the physical parameters of the body T are not connected to the others by the equation (1.2).

Thus, only geological bodies with established dependences between various physical properties can be targets for the joint interpretation of multimethod geophysical data.

The only exception from this rule is the case in which different geophysical methods are used to establish the distribution of one physical parameter, e.g. the combined use of DC and EM sounding to investigate the electrical resistivity of the Earth.

From (1.1) and (1.2) it follows that the geophysical fields $U_1(\mathbf{r}), U_2(\mathbf{r}),..., U_k(\mathbf{r})$ caused by the body T and observed in the area Q are also connected, by a relationship:

$$Y[U_1(\mathbf{r}), U_2(\mathbf{r}),..., U_k(\mathbf{r})] = 0 \, . \tag{1.4}$$

The classical example of such a relationship is the Poisson relation that connects the gravity and magnetic potentials of a body with uniform density and magnetisation.

Since the problem (1.3) cannot be solved analytically for an arbitrary distribution of physical parameters within the body T, numerical methods have to be employed for this purpose. One possible way is to approximate the unknown geometry of the body T by some model described via geometrical parameters $\mathbf{g} = (g_1, g_2,..., g_s)$ that together with the distribution of physical parameters $\beta_i(\mathbf{t}), i = 1,2,..., k$ form the parameter vector $\mathbf{m} = (m_1, m_2,..., m_p)$ common to all the data sets under consideration. Note that if equation (1.2) connecting the physical parameters is known, then some of $\beta_i(\mathbf{t})$ can be expressed via other parameters and the number of unknown parameters of the body T can be reduced. It is possible to find an estimate for the vector $\mathbf{m}$ as a solution to the problem of *joint inversion* that is implemented by minimising the differences between all the

geophysical fields observed in the area Q and corresponding values calculated from the parameter vector $\mathbf{m}$:

$$\begin{cases} \text{minimise } J_1[U_1^{obs}(\mathbf{r}), F_1(\mathbf{m},\mathbf{r})], \\ \quad\quad\quad ... \quad\quad\quad\quad\quad\quad \mathbf{m} \in M,\ \mathbf{r} \in Q, \\ \text{minimise } J_k[U_k^{obs}(\mathbf{r}), F_k(\mathbf{m},\mathbf{r})], \end{cases} \tag{1.5}$$

where M is a set of feasible values of the vector $\mathbf{m}$ and $J_i[U_i^{obs}(\mathbf{r}), F_i(\mathbf{m},\mathbf{r})], i = 1,...,k$ is a function characterising the discrepancy between the observed values for the ith geophysical field and the corresponding theoretical response calculated from the parameter vector $\mathbf{m}$.

One more problem in obtaining parameters for the multiresponse geophysical body is that the field $U_i(\mathbf{r})$ measured in the observation area Q is a combined effect of the body T and other field sources formed as a result of geological and physical processes that did not necessary affect all the physical properties simultaneously. These field sources do not have physical parameters connected by equation (1.2) and can be called *single-response bodies* with respect to the body T. If the target of data interpretation is to find parameters of the body T, then the effect of single-response field sources can be regarded as noise that may be of the same order as the field produced by the body T. The effect of single-response bodies can be estimated if equation (1.2) or its approximation is known. In this case one of the physical parameters can be expressed via others:

$$\beta_i(\mathbf{t}) = V[\beta_1(\mathbf{t}),..., \beta_{i-1}(\mathbf{t}), \beta_{i+1}(\mathbf{t}),..., \beta_k(\mathbf{t})], \tag{1.6}$$

where $V[\cdot]$ denotes an operator. The equation (1.6) can then be substituted into the corresponding equation (1.3) for the ith field:

$$U_i(\mathbf{r}) = F_i\{V[\beta_1(\mathbf{t}),..., \beta_{i-1}(\mathbf{t}), \beta_{i+1}(\mathbf{t}),..., \beta_k(\mathbf{t})]\}. \tag{1.7}$$

Parameters $\beta_j(\mathbf{t}), j = 1,2,...,k, j \neq i$ can in turn be expressed from observed geophysical fields $U_j^{obs}(\mathbf{r}), j = 1,2,...,k, j \neq i$, assuming that the corresponding inverse operators F_j^{-1} are known for F_j in (1.3):

$$\beta_j(\mathbf{t}) = F_j^{-1}[U_j^{obs}(\mathbf{r})], j = 1,2,...,k, j \neq i. \tag{1.8}$$

Equation (1.8) can then be substituted into (1.7), giving an expression for the effect of the multiresponse body T in the field $U_i^{obs}(\mathbf{r})$. The difference between $U_i^{obs}(\mathbf{r})$ and $U_i(\mathbf{r})$ calculated from eq. (1.7-1.8) gives an expression for the effect of single-response bodies with an anomalous distribution of the single parameter $\beta_i(t)$:

$$\Delta U_i(\mathbf{r}) = U_i^{obs}(\mathbf{r}) - F_i\{V[\beta_1(\mathbf{t}),..., \beta_{i-1}(\mathbf{t}), \beta_{i+1}(\mathbf{t}),..., \beta_k(\mathbf{t})]\}. \tag{1.9}$$

The effect of single-response bodies, described by eq. (1.9), can be the object of a separate interpretation.

The approach described by equations (1.6 -1.9) is difficult to implement analytically because precise analytical expressions for equation (1.6) or for the inverse operators (1.8) very seldom exist. As demonstrated in Papers II and III, it is possible in some cases to approximate the equation (1.6) by simpler expressions depending on *a-priori* information. The method of joint interpretation of seismic and gravity data discussed in Chapter 2 and presented in Papers II and III is based on approximation of the relationship between density and seismic velocity, which is constructed from *a-priori* data.

The effect of single-response bodies within the area under study can be as large as the effect of the multiresponse body *T,* and cannot always be regarded as non-informative noise. One of the purposes of the present investigation was to demonstrate that bodies for which the established relationship between physical properties is violated could be targets of separate interpretations. It is demonstrated in Chapter 2 that the areas in the crust where the well-known dependence between seismic velocity and density is violated are often associated with important tectonic boundaries, and it is shown in Chapter 3 how the concept of a *multiscale medium* can be used to separate the effect of common factors (pores and fractures in the upper crust) affecting both seismic velocity and electrical conductivity from a factor that affects only velocity (rock mineral composition).

1.2 Formulation of the problem of joint inversion of multimethod geophysical data

The traditional geophysical inverse problem was formulated for one separate experimental data set by Goltsman (1982) and Tarantola (1987) as follows.

Let T be a geological object (body) described by a parameter vector $\mathbf{m}=[m_1,m_2,\ldots,m_n]\in M$, where M is a set of feasible solutions defined in the parameter space R^n. Let us assume that some *a-priori* information about the object T was obtained from previous independent experimental observations and can be expressed as a probability density function (PDF) $p(\mathbf{m})$. The new information about the geological object T, i.e. about the vector **m,** can be obtained by inverting data from a geophysical experiment that form the observation vector $\mathbf{d_{obs}} = (d_1, d_2, ..., d_Q)$. The *a-posteriori* conditional probability density of the vector **m** after the experiment is (Tarantola 1987):

$$p(\mathbf{m} \mid \mathbf{d}_{\mathrm{obs}}) = p(\mathbf{m})\, p(\mathbf{d_{obs}} \mid \mathbf{m})/p(\mathbf{d_{obs}}), \tag{1.10}$$

where $p(\mathbf{m})$ is the *a-priori* PDF of **m,** $p(\mathbf{d_{obs}} \mid \mathbf{m})$ is the conditional probability density of the experimental data (probability of obtaining certain experimental data for given values of model parameters), and $p(\mathbf{d_{obs}})$ is the marginal probability density of the experimental data. As $p(\mathbf{d_{obs}})$ does not depend on **m,** it can be assumed to be equal to some b=const after the experimental data have been obtained. The maximum of the conditional probability density function (1.10) gives an estimate for the parameter vector **m** that is called the maximum likelihood solution**:**

$$\mathbf{m}^* = \max_{\mathbf{m}}(p(\mathbf{m} \mid \mathbf{d_{obs}})) = \max_{\mathbf{m}}(p(\mathbf{m})\, p(\mathbf{d_{obs}} \mid \mathbf{m})). \tag{1.11}$$

The conditional probability density of the experimental data in (1.10) depends on an operator g that allows theoretical values for observable geophysical fields to be calculated from known values of model parameters subject to a certain assumption regarding the signal-noise model M, i.e.

$$\mathbf{d_{teor}} = g(\mathbf{m}), \qquad \mathbf{d_{obs}} = M(g(\mathbf{m}), \mathbf{n}) . \tag{1.12}$$

Depending on the type of PDF in (1.10-1.11), one can obtain different formulations of geophysical inverse problems. Goltsman (1982) and Tarantola (1987) demonstrated that if the PDF $p(\mathbf{m})$ and $p(\mathbf{d_{obs}} \mid \mathbf{m})$ in (1.11) are generalised Gaussian of order Q and n, respectively, where $1 \le Q < \infty, 1 \le n < \infty$, the problem (1.11) is equivalent to minimisation of the objective function

$$L(\mathbf{m}) = \| g(\mathbf{m}) - \mathbf{d_{obs}} \|_{l_Q} + \| \mathbf{m} - \mathbf{m_0} \|_{l_n} , \tag{1.13}$$

where $\|g(\mathbf{m}) - \mathbf{d_{obs}}\|_{l_Q}$ denotes the difference between the theoretically calculated and observed geophysical fields in the weighted l_Q norm and $\|\mathbf{m} - \mathbf{m_0}\|_{l_n}$ denotes the difference between $\mathbf{m}$ and the *a-priori* model $\mathbf{m_0}$ in the weighted l_n norm.

Maximisation of the PDF (1.10) or minimisation of (1.13) in real geophysical inverse problems requires the application of numerical optimisation techniques. Although it is often assumed that the PDFs in (1.10) and (1.13) are defined on an infinite set and are continuous, we have to deal in practice with a finite data set measured by a digital device of limited accuracy. We also have to use computers to calculate the theoretical response of the vector of the model parameters and perform optimisation of (1.10) or (1.13) numerically. As a result, one has in reality to optimise a discrete function that can be calculated with limited accuracy and is defined on a finite and discrete set of feasible values of the parameter vector **m.**

The other problem is non-uniqueness, which can be called a fundamental property of geophysical inverse problems. Two types of non-uniqueness can be distinguished. The first type can be called theoretical non-uniqueness, which arises due to the existence of equivalent field sources that cause the same response in the observation area. There exist certain classes of field sources for which it is possible to prove a "uniqueness theorem", i.e. to establish theoretically that the inverse problem has a unique solution for perfect, error-free data (uniqueness proof). The classical example of a uniqueness theorem is Novikov's lemma (Novikov, 1937) concerning conditions for the uniqueness of the solution of an inverse gravimetric problem in a three-dimensional space. Unfortunately, one very often finds that such classes of field sources cannot be used to describe real geological objects.

The second type of non-uniqueness originates from the fact that there exist sources producing different fields in the observation area but giving equal, or near equal values of the PDF (1.10) or objective function (1.13), due to the finite set of experimental data and the presence of noise (Parker, 1977). This type of non-uniqueness can never be avoided, as one always has to deal with finite, noisy experimental data.

Due to the non-uniqueness of the real geophysical inverse problem, there exists a set B of vectors $\mathbf{m}$ that maximise the *a-posteriori* PDF (1.10) with finite accuracy $\delta > 0$, i.e.

$B = \{\mathbf{m} \,\|\, \|p(\mathbf{m} \mid \mathbf{d_{obs}}) - p(\mathbf{m}^{true} \mid \mathbf{d_{obs}})\| < \delta\}$, where $\mathbf{m}^{true}$ corresponds to the "true" maximum likelihood point of theoretical distribution. Since this set can contain geologically meaningless solutions as well, the proper solution to a problem of geophysical data inversion can be found only after analysis of all possible solutions from this set. The model that agrees with the information obtained by the other methods can then be accepted as the final solution. If the geological body T can be described by a parameter vector $\mathbf{m}$ that is common to all methods**,** joint inversion of the multimethod geophysical data can be applied to obtain a model that agrees with all the data sets.

Suppose that we have the results of k independent geophysical experiments carried out in the same observation area, i.e. a set of observation vectors $\mathbf{d}_{obs} = \{\mathbf{d}^1_{obs}, \mathbf{d}^2_{obs}, ..., \mathbf{d}^k_{obs}\}$. If all the k geophysical experiments are independent, then all $\mathbf{d}^i_{obs}, i = 1,...,k$ can be regarded as independent random variables and the *a-posteriori* PDF of vector $\mathbf{m}$ after all the sets of experimental data were obtained can be written as (Goltsman, 1982)

$$p(\mathbf{m} \mid \mathbf{d}^1_{\text{obs}}, \mathbf{d}^2_{\text{obs}}, ..., \mathbf{d}^k_{\text{obs}}) = bp(\mathbf{m})\, p(\mathbf{d}^1_{\mathbf{obs}} \mid \mathbf{m}) p(\mathbf{d}^2_{\mathbf{obs}} \mid \mathbf{m}) ... p(\mathbf{d}^k_{\mathbf{obs}} \mid \mathbf{m}) \,, \qquad (1.14)$$

where b=const represents the marginal PDF of the experimental data that does not depend on $\mathbf{m}$.

Maximisation of (1.14) gives an estimate for the parameter vector $\mathbf{m}^*$, i.e. a solution to the problem of joint inversion.

Equation (1.14) is the most common formulation of the problem of joint inversion of multimethod geophysical data, as already proposed by Goltsman and Kalinina in the 1960s. It follows directly from the multiplication theorem for the probabilities of independent events (Pugachev, 1965). Namely, if two events A and B are independent, then the probability of their combined appearance $P(A \cap B) = P(A)P(B)$.

As already mentioned, it is usually assumed that all the PDFs $p(\mathbf{m})$ and $p(\mathbf{d}^i_{\mathbf{obs}} \mid \mathbf{m}), i = 1,...,k$ in (1.14) are generalised Gaussian with diagonal covariance matrices of order Q_i and n, respectively, where $1 \le Q_i < \infty, 1 \le n < \infty$. In this case maximisation of (1.14) is equivalent to minimisation of the objective function

$$L(\mathbf{m}) = \sum_{i=1}^{k} \left\| g_i(\mathbf{m}) - \mathbf{d}^i_{\mathbf{obs}} \right\|_{l_{Q_i}} + \left\| \mathbf{m} - \mathbf{m_0} \right\|_{l_n} , \qquad (1.15)$$

where $\left\| g_i(\mathbf{m}) - \mathbf{d}^i_{\mathbf{obs}} \right\|_{l_{Q_i}}$ denotes the difference between the theoretically calculated and observed geophysical fields in a weighted l_{Q_i} norm and $\left\| \mathbf{m} - \mathbf{m_0} \right\|_{l_n}$ denotes the difference between $\mathbf{m}$ and the *a-priori* model $\mathbf{m_0}$ in a weighted l_n norm. If all the PDFs in (1.14) are Gaussian of order 2 and the operators $g_i(\mathbf{m})$ are linear, problem (1.15) becomes a linear one. In this case the resolution analysis can be performed and the inversion efficiency estimated on the basis of covariance matrices (Tarantola, 1987, Menke, 1989, Hjelt, 1992). This analysis is not valid in the case of non-linear problems, due to the multimodality of the PDF (1.14) and the objective function (1.15), so that the efficiency of the joint inversion of multimethod geophysical data is analysed here using the definition of ε*-entropy* (Kolmogorov and Tichomirov, 1959). As can be seen from the following definitions, ε -entropy makes it possible to analyse non-uniqueness in real geophysical inverse problems which have finite accuracy and a finite set of feasible solutions. By means of ε -entropy it is possible to demonstrate that joint inversion of

multimethod geophysical data can reduce non-uniqueness in a real geophysical inverse problem.

Let A be a non-empty set in a space R^k with a norm denoted by $\| \ \|$, and let ε be a positive number. Then the following subsets of set A can be introduced (Kolmogorov and Tichomirov, 1959):

Definition 1.1. The system γ of sets $U_j \subset R^k, j \in J$ is the *ε-partition* of the set $A \subset R^k$ if the diameter $d(U_j)$ of any $U_j \subset \gamma$ does not exceed 2ε and $A \subseteq \bigcup_{U_j \in \gamma} U_j$.

Definition 1.2. The set $U \subset R^k$ is the *ε-net* of the set A if for every point $\mathbf{a} \in A$ there exists a point $\mathbf{u} \in U$ such that $\|\mathbf{a} - \mathbf{u}\| < \varepsilon$.

The following statements can be proved for any $A \subset R^k$:

Theorem 1.1. (for proof, see Kolmogorov and Tichomirov (1959)

a) For any $\varepsilon > 0$ there exists a finite ε-partition of the set A.

b) For any $\varepsilon > 0$ there exists a finite ε-net of the set A.

Then the following characteristics of the set A can be defined.

Definition 1.3. Let $\mathrm{N}_\varepsilon(A)$ be the minimal number of sets in the ε-partition of the set A and $\mathrm{N}_\varepsilon^{R^k}(A)$ be the minimal number of points in the ε-net of the set A. Then the function $H_\varepsilon(A) = \log_2 \mathrm{N}_\varepsilon(A)$ is the *minimal entropy of the set A* or *ε-entropy of the set A*. The function $H_\varepsilon^{R^k}(A) = \log_2 \mathrm{N}_\varepsilon^{R^k}(A)$ is the *ε-entropy of the set A with respect to* R^k.

If the space R^k is centred (i.e. if $\forall U \subset R^k$ with diameter $d(U) = 2r$ $\exists \mathbf{x^0} \in U \mid \left\|\mathbf{x} - \mathbf{x^0}\right\| < r \ \forall \mathbf{x} \in U$), then $\mathrm{N}_\varepsilon(A) = \mathrm{N}_\varepsilon^{R^k}(A)$ and $H_\varepsilon(A) = H_\varepsilon^{R_k}(A)$.

The difference between the Shannon entropy used in classical information theory and minimal ε-entropy is clear from the above definition. The entropy of the PDF $p(\mathbf{m}), \mathbf{m} \in A$ is an informational measure of the uncertainty of the random vector **m,** defined as (Shannon, 1948)

$$H = -\int_A p(\mathbf{m}) \log p(\mathbf{m}) d\mathbf{m} \,. \tag{1.16}$$

This is a function of the probability density function, i.e. of a measure of the set A. On the other hand, the ε-entropy can be regarded as a function of the set A and ε. It does not depend on any probability measure defined on A, i.e. it is similar to the Hartley function introduced as a measure of the uncertainty of finite sets (Hartley, 1928):

$$U(A) = c \cdot \log_b |A|, \tag{1.17}$$

where $b>1$ and $c>0$ are positive constants and $|A|$ denotes the cardinality of a finite non-empty set A.

If A is a set of feasible solutions to the inverse problem (1.11) in which two vectors **m** and $\mathbf{m}^1$ are distinguishable if $\left\|\mathbf{m} - \mathbf{m^1}\right\| > \varepsilon$, where ε is a positive number, then $\mathrm{N}_\varepsilon(A)$ is the minimal number of solutions that can be recognised in the set A with the accuracy ε, and $H_\varepsilon(A)$ is the measure of uncertainty of the set A, which can also be used to estimate the non-uniqueness of the inverse problem under consideration.

The minimal value of ε for any geophysical inverse problem depends on many factors, the main ones being model parameterisation, the correlation between the model parameters,

the accuracy of forward problem calculation and noise in the observed data. Two parameter vectors $\mathbf{m}$ and $\mathbf{m}^1$ can be distinguished in the parameter space of a geophysical inverse problem (1.11) if the following condition is satisfied:

$$\left\|\mathbf{m}-\mathbf{m}^{\mathbf{1}}\right\| > \varepsilon \Rightarrow \left\| p(\mathbf{m} \mid \mathbf{d}_{\mathbf{obs}}) - p(\mathbf{m}^1 \mid \mathbf{d}_{\mathbf{obs}}) \right\| > \delta \,, \tag{1.18}$$

where δ is the lower limit of accuracy that can be obtained in the calculation of function (1.11). The corresponding value of ε defines the *lower limit of resolution* of the inverse problem under consideration, in the sense that if $\left\|\mathbf{m}-\mathbf{m}^{\mathbf{1}}\right\| < \varepsilon$, the difference between $p(\mathbf{m} \mid \mathbf{d}_{\mathbf{obs}})$ and $p(\mathbf{m}^1 \mid \mathbf{d}_{\mathbf{obs}})$ becomes insignificant.

Kolmogorov and Tichomirov (1959) demonstrated that if the two sets A and B belong to the same set C and are independent, i.e. $A \subset C$, $B \subset C$ and $A \cap B = \varnothing$, then the ε-entropy of the set $C = A \cup B$ is

$$H_\varepsilon(C) = H_\varepsilon(A) + H_\varepsilon(B) \,. \tag{1.19}$$

From (1.19) it follows that if $A \subseteq B$, then $H_\varepsilon(A) \le H_\varepsilon(B)$.

The *a-posteriori* PDF of vector $\mathbf{m}$ in the case in which only data from the ith experiment are known is

$$p(\mathbf{m} \mid \mathbf{d}^{\mathrm{i}}_{obs}) = b\, p(\mathbf{m})\, p(\mathbf{d}^{i}_{obs} \mid \mathbf{m}),\ i = 1,2,...,k \,. \tag{1.20}$$

Let $B_i \subset A$ be a set of estimates $\mathbf{m}^*$ that maximise the ith *a-posteriori* PDF in (1.20) with the accuracy δ_i, i.e.

$$B_i = \left\{ \mathbf{m}^* \mid \mathbf{m}^* \in A, \left\| p(\mathbf{m}^* \mid \mathbf{d}^i_{obs}) - p(\mathbf{m}^{true} \mid \mathbf{d}^i_{obs}) \right\|_{l_1} < \delta_i \right\} , \tag{1.21}$$

where $\mathbf{m}^{true}$ is the "true" maximum likelihood point of PDF (1.20), which cannot be achieved in reality due to discreteness and numerical errors.

Then the uncertainty of the random vector $\mathbf{m}$ after separate interpretation of the data from the ith experiment is estimated by the entropy $H_\varepsilon(B_i \mid \mathbf{d}^i_{\mathbf{obs}})$, where ε is the lower limit of resolution of the ith method. As $B_i \subset A$, then $H_\varepsilon(B_i \mid \mathbf{d}^i_{\mathbf{obs}}) \le H_\varepsilon(A)$.

The set of estimates of vector $\mathbf{m}^*$ obtained from joint interpretation of the data from all k experiments must maximise the joint *a-posteriori* PDF (1.14). Consider the set of vectors $\mathbf{m}^1$ that simultaneously maximises all the *a-posteriori* PDFs (1.20) with the accuracy δ_i, $i=1,...,k$. i.e. $D = \bigcap_{i=1}^{k} B_i$ and $D \subseteq B_i$. Obviously, if a vector $\mathbf{m}^1 \in D$, then it also maximises the joint *a-posteriori* PDF (1.14), i.e. D is the set of solutions to the problem of joint interpretation. Note that in the case in which $D = \varnothing$ it is always possible to increase some values of δ_i and obtain $D \neq \varnothing$. The large values of δ_i in this case indicate that either the model parameterisation is not correct or the PDFs making up the joint PDF (1.14) have not been selected properly.

The values for the lower resolution limits ε_i are usually different for each separate method. However, it follows from Theorem 1.1 that for every set B_i it is possible to define the ε-partition with ε corresponding to the method with the smallest value for the lower limit of resolution. This means that in principle it is always possible to construct a procedure for joint inversion of multimethod geophysical data with a lower limit of

resolution equal to that of the method with the best resolution. The other conclusion is that the lower limit of resolution of any procedure of joint inversion of multimethod geophysical data cannot be less than that of the method with the best resolution.
In this case the entropy of vector **m** after joint interpretation of the data from k independent experiments is $H_\varepsilon(D \mid \mathbf{d}^1_{\mathbf{obs}}, \mathbf{d}^2_{\mathbf{obs}}, ..., \mathbf{d}^k_{\mathbf{obs}})$. Then it follows from $D \subseteq B_i,\ i = 1,2,...,k$ that

$$H_\varepsilon(D \mid \mathbf{d}^1_{\mathbf{obs}}, \mathbf{d}^2_{\mathbf{obs}}, ..., \mathbf{d}^k_{obs}) \le H_\varepsilon(B_i \mid \mathbf{d}^i_{\mathbf{obs}}) \le H_\varepsilon(A)\,. \tag{1.22}$$

Equality of $H_\varepsilon(D \mid \mathbf{d}^1_{\mathbf{obs}}, \mathbf{d}^2_{\mathbf{obs}}, ..., \mathbf{d}^k_{\mathbf{obs}})$ and $H_\varepsilon(B_i \mid \mathbf{d}^i_{\mathbf{obs}})$ in (1.22) is achieved only if the sets D and B_i coincide. Inequality (1.22) means that joint inversion of data from several geophysical methods can reduce non-uniqueness in real geophysical inverse problems.
The following conclusions can be reached from the above analysis:

a) The lower limit of resolution of any procedure of joint inversion applied to multimethod geophysical data corresponds to the value for the method with the best resolution, but cannot be less than this.
b) As joint inversion of multimethod geophysical data can reduce non-uniqueness in real geophysical inverse problems, the solution obtained from such data is better then that obtained from separate inversion of data from only one method.

It is important to note that these conclusions do not depend on the type of PDF in a concrete joint inversion algorithm or on the linearity of the problem.

1.3 Joint inversion of multimethod geophysical data as a multiobjective optimisation problem (MOP)

1.3.1 Formulation of the problem

As demonstrated in the previous section, the problem of joint inversion of multimethod geophysical data can be formulated as a problem of simultaneous maximisation of all conditional *a-posteriori* PDFs of the vector of the model parameters **m** that are included in equation (1.14), i.e. it can be formulated as a problem of the kind (1.5) with the functions $J_i[U_i^{obs}(\mathbf{r}), F_i(\mathbf{m}, \mathbf{r})] = -p(\mathbf{m} \mid \mathbf{d}_{obs}), i = 1,..., k$.

This is *a multiobjective optimization problem (MOP).* The theory of multiobjective optimisation is a part of optimisation theory that deals with the case when the objective function is a vector. It is assumed in the present work that the inverse problem is always formulated as a minimisation problem, because any maximisation problem can be transformed into a minimisation one. The MOP can then be formulated as follows:

Minimise the vector objective function $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}),..., f_k(\mathbf{x}))$ that assumes values in R^k subject to $\mathbf{x} = (x_1, x_2,..., x_n) \in X \subseteq R^n$, where X is a set of feasible solutions to the MOP, i.e.

$$\text{minimise } \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}),..., f_k(\mathbf{x})), \quad \mathbf{x} \in X, \tag{1.23}$$

where the set of feasible solutions *X* is defined by a system of inequality constraints.

The vector objective function maps the set of feasible solutions *X* from the *solution space* R^n to the *objective space* R^k. If all the functions $f_i(\mathbf{x}), i = 1,...,k$ are linear and the set of feasible solutions *X* is defined by a system of linear inequalities, then the problem (1.23) is called a *multiobjective linear optimisation problem (MOLP).*

If some of the functions $f_i(\mathbf{x}), i = 1,...,k$ are non-linear and/or the set of feasible solutions *X* is defined by a system of constraints that is composed of both non-linear and linear inequalities, then the problem (1.23) is called a *multiobjective non-linear optimisation problem (MONLP).*

A non-linear multiobjective problem is called a *convex MONLP* if all $f_i(\mathbf{x}), i = 1,...,k$ as well as the inequalities defining the set of feasible solutions *X* are convex functions.

The following definitions are a generalisation of the definition of the minimum of a scalar function to the case of a vector objective function (Sakawa, 1993).

Definition 1.4 (complete optimal solution): x^* is said to be a complete optimal solution of a MOP with the vector objective function $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}),..., f_k(\mathbf{x}))$ if and only if there exists $\mathbf{x}^* \in X$ such that $f_i(\mathbf{x}^*) \leq f_i(\mathbf{x}),\ i = 1,...,k$ for all $\mathbf{x} \in X$.

Definition 1.5 (weak Pareto optimal solution): $\mathbf{x}^*$ is said to be a weak Pareto optimal solution of a MOP with the vector objective function $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}),..., f_k(\mathbf{x}))$ if and only if no other $\mathbf{x} \in X$ exists such that $f_i(\mathbf{x}) < f_i(\mathbf{x}^*),\ i = 1,...,k$ for all *i.*

Definition 1.6 (Pareto optimal solution): $\mathbf{x}^*$ is said to be a Pareto optimal solution of a MOP with the vector objective function $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}),..., f_k(\mathbf{x}))$ if and only if no other $\mathbf{x} \in X$ exists such that $f_i(\mathbf{x}) \leq f_i(\mathbf{x}^*),\ i = 1,...,k$ for all *i* and $f_j(\mathbf{x}) \neq f_j(\mathbf{x}^*)$ for at least one *j*.

Let X^{CO}, X^P, X^{WP} denote sets of complete optimal, Pareto optimal and weak Pareto optimal solutions of a MOP, respectively. It then follows from Definitions 1.4-1.6 that if a complete optimal solution to the MOP exists, $X^{CO} \subseteq X^P \subseteq X^{WP}$.

The objective functions in a real MOP usually contradict each other, and no complete optimal solution may exist. On the other hand, the sets of Pareto optimal solutions (Pareto set) and weak Pareto optimal solutions (weak Pareto set) always exist. Consequently, the methods of multiobjective optimisation aim at finding a set of Pareto optimal solutions to the MOP, or at least a set of weak Pareto optimal solutions. It is seen from Definitions 1.5-1.6 that any Pareto optimal solution is a trade-off solution: it is not possible to improve it by reducing one objective function without increasing the others. Since the number of trade-off possibilities in a MOP can be fairly large, the set of Pareto optimal solutions to a real MOP may contain a variety of solutions with different trade-off rates. Obviously, the larger the number of objectives of the MOP, the larger will be the number of possible trade-off variants. Selection of the final solutions from the Pareto set requires

analysis of the information on the trade-off rate between various objectives and is a necessary stage in the solution of the MOP.

Consider the set of vectors $\mathbf{m}^*$ that simultaneously maximise all PDFs in (1.14), i.e. $D = \bigcap_{i=1}^{k} B_i$. In accordance with Definition 1.4, this is a complete set of optimal solutions to a multiobjective problem (1.5) in which the *a-posteriori* PDFs (1.20) are used as objective functions. Although it is not correct to say that the joint inversion of multimethod geophysical data is a problem with highly contradictory objectives, errors in the experimental data and approximation of the real medium by a parameter vector can mean that no complete optimal solution exists, i.e. $D = \varnothing$. As has already been mentioned in section 1.2, it is possible in this case to require that some or all of the *a-posteriori* PDFs (1.20) should be as close to the optimum value as possible, i.e. to allow a trade-off between PDFs. This can be done by the following very simple procedure, for example:

Step 1. Find the set $B_1 \subset A$ of estimates $\mathbf{m}^*$ that minimise the 1st objective function in (1.5).

Step 2. Calculate the values for the second objective function at each point in the set B_1.

Step 3. Find the set $B_2 \subset B_1$ of estimates $\mathbf{m}^1$ with a minimum value for the 2nd objective function.

Step 4. Calculate the values of the 3rd objective function at each point in the set $B_2 \subset B_1$.

Step 5. Find the set $B_3 \subset B_2 \subset B_1$ of estimates $\mathbf{m}^2$ with the minimum value for the 3rd objective function.

Step 6. Calculate the values of the 4th objective function at each point in the set $B_3 \subset B_2 \subset B_1$, etc.

The process stops when the entire vector of objective functions in (1.5) is being considered, or when only one point remains in the set B_i. The points in the final set $B_k \subset B_{k-1} \subset ... \subset B_1$ are Pareto optimal in accordance with Definition 1.6. Note that this set does not contain all the possible trade-off solutions and cannot be accepted as the best way to solve the MOP.

Among the variety of multiobjective optimisation techniques there exist a group of *scalarization methods* which allow transformation of the MOP into a one-objective problem. These methods usually aim at finding the solution in the Pareto set in accordance with the *a-priori* defined trade-off rate between the various components of the vector objective function. Only the main methods for scalarization of multiobjective problems are analysed here, as some of them are used explicitly in the practice of geophysical data inversion.

1.3.2 The weighting method of scalarization for a MOP

The weighting method, originally proposed by Kuhn and Tucker (1951), involves solving the following weighting problem instead of (1.23):

$$\text{minimise} \quad f(\mathbf{x}) = \sum_{i=1}^{k} w_i f_i(\mathbf{x})$$
$$\text{subject to } \mathbf{x} \in X \,, \tag{1.24}$$

where $\mathbf{w} = (w_1,...,w_k) \geq 0$ is the vector of weighting coefficients assigned to the components of the vector objective function prior to solution of the problem. The weighting coefficients characterise the rate of trade-off between different components of the vector objective function; i.e. the selection of a solution from the Pareto set depends entirely on the selected vector **w**.

In the case of MOLP the relationship between the optimal solution to the problem (1.24) and Pareto optimality is given by the following two theorems, the proofs of which, as for other theorems concerning the Pareto optimality of solutions to MOPs, can be found in the book by Sakawa (1993):

Theorem 1.2. If $\mathbf{x}^* \in X$ is an optimal solution to the weighting problem for some $\mathbf{w} = (w_1,...,w_k) \geq 0$, then $\mathbf{x}^*$ is a Pareto optimal solution to the MOLP.

Theorem 1.3. If $\mathbf{x}^* \in X$ is a Pareto optimal solution to the MOLP, then $\mathbf{x}^*$ is an optimal solution to the weighting problem for some $\mathbf{w} = (w_1,...,w_k) \geq 0$.

Note that Theorem 1.2 can also be proved for MONLP, whereas Theorem 1.3 can be proved for arbitrary $\mathbf{w} = (w_1,...,w_k) \geq 0$ only in the case of MOLP and convex MONLP. This means that in the case of non-linear problems there can be Pareto optimal solutions for which no corresponding weighting problem exists. Podinovsky and Nogin (1982) demonstrated that in the case of a non-convex MONLP there can be vectors $\mathbf{w} = (w_1,...,w_k) \geq 0$ for which the minimum of (1.24) does not exist within the set of feasible solutions *M*. This property of MONLP must be taken into consideration when employing the weighting method to solve non-linear problems involving joint inversion of multimethod geophysical data.

The weighting method is one of the most popular ways of treating the problem of joint inversion in the practice of geophysical data interpretation. In the weighting method the functions $J_i[U_i^{obs}(\mathbf{r}), F_i(\mathbf{m},\mathbf{r})], i = 1,...,k$ are normalised and combined into a sum, the minimisation of which gives an estimate for the parameter vector **m**:

$$\mathbf{m}^* = \min_{\mathbf{m}} \sum_{i=1}^{k} w_i J_i(U_i^{obs}(\mathbf{r}), F_i(\mathbf{m},\mathbf{r})) \,. \tag{1.25}$$

The problem of joint inversion of multimethod geophysical data, formulated as (1.15), i.e. under the assumption that all *a-posteriori* PDFs in (1.14) are generalised Gaussian, can also be regarded as the weighted sum of the objective functions of the following MOP:

$$\text{minimise } \mathbf{L}(\mathbf{m}) = (L_1(\mathbf{m}), L_2(\mathbf{m}),..., L_k(\mathbf{m}), L_{k+1}(\mathbf{m})) \text{ where}$$

$$
\begin{aligned}
&L_1(\mathbf{m}) = \left\| g_1(\mathbf{m}) - \mathbf{d}^1_{\mathbf{obs}} \right\|_{l_{Q_1}}, \\
&L_2(\mathbf{m}) = \left\| g_2(\mathbf{m}) - \mathbf{d}^2_{\mathbf{obs}} \right\|_{l_{Q2}}, \\
&\ldots \\
&L_k(\mathbf{m}) = \left\| g_k(\mathbf{m}) - \mathbf{d}^1_{\mathbf{obs}} \right\|_{l_{Q_k}}, \\
&L_{k+1}(\mathbf{m}) = \left\| \mathbf{m} - \mathbf{m}_0 \right\|_{l_n}, \\
&\text{subject to } \mathbf{m} \in M.
\end{aligned}
\tag{1.26}
$$

In (1.26) $\left\| g_i(\mathbf{m}) - \mathbf{d}^i_{\mathbf{obs}} \right\|_{l_{Q_i}}$ denotes the difference between the theoretically calculated and observed values for the ith geophysical field in a weighted l_{Q_i} norm and $\left\| \mathbf{m} - \mathbf{m_0} \right\|_{l_n}$ the difference between $\mathbf{m}$ and the *a-priori* model $\mathbf{m_0}$ in the weighted l_n norm.

In the case of linear joint inversion and convex non-linear joint inversion problems the estimate given by minimisation of (1.15) is a Pareto optimal solution to the problem (1.26). In the case of non-convex problems the minimum of (1.15) does not necessary exist, and numerical optimisation algorithms can fail.

A multiobjective interpretation can also be given for the method of solving ill-posed problems known as Tikhonov regularisation (Tikhonov and Arsenin, 1977), in which an estimate for the parameter vector $\mathbf{m}$ is obtained as a solution to the minimisation problem

$$
\mathbf{m}^* = \min_{\mathbf{m}} [J(g(\mathbf{m}), \mathbf{d}_{\mathbf{obs}}) + \alpha \Omega(\mathbf{m})], \alpha > 0, \tag{1.27}
$$

where $J(g(\mathbf{m}), \mathbf{d}_{\mathbf{obs}})$ is a function describing the difference between the experimental data and a calculated model response (the difference in the l_2 norm is usually used), $\Omega(\mathbf{m})$ is a stabilising function defined on the set of feasible solutions M and $\alpha > 0$ is a regularisation parameter laid down *a-priori*. As demonstrated by Sobol (1985), if $\mathbf{m}_\alpha$ is the solution to the problem (1.27) for some $\alpha>0$, then $\mathbf{m}_\alpha \in C$, where C is the Pareto set of the following two-objective problem:

$$
\begin{aligned}
&\text{minimise } \mathbf{f}(\mathbf{m}) = (f_1(\mathbf{m}), f_2(\mathbf{m})) \text{ where} \\
&f_1(\mathbf{m}) = J[g(\mathbf{m}), \mathbf{d}_{\mathbf{obs}}], \\
&f_2(\mathbf{m}) = \Omega(\mathbf{m}), \\
&\text{subject to } \mathbf{m} \in M.
\end{aligned}
\tag{1.28}
$$

In this case problem (1.27) can be regarded as scalarization of a two-objective problem by a weighted sum of objective functions.

It is known that Tikhonov regularisation was originally proposed for the case in which $g(\mathbf{m})$ is a linear, finite operator, and experience with its application to problems with a non-linear operator $g(\mathbf{m})$ has demonstrated that it is much less effective in that case. The reason for this is that in the case of non-linear, non-convex components of the vector

objective function, the minimum of the weighted sum does not exist for any vector of weights, i.e. for any $\alpha > 0$. Consequently use of the Tikhonov regularisation technique cannot be recommended in the case of non-linear inverse problems for which convexity is not guaranteed.

1.3.3 The constraint method of scalarization of a MOP

The constraint method (Haimes et al., 1971) is not so sensitive to non-convexity as the weighting method considered above. In this method the MOP is formulated as a constraint problem by taking one component of the vector objective function as the main objective and constraining the other components by means of inequalities, i.e.

$$\begin{aligned} &\text{minimise } f_i(\mathbf{x}) \\ &\text{subject to } f_j(\mathbf{x}) \le \varepsilon_j, \ \ j = 1,...,k, \ \ j \ne i, \\ &\qquad \mathbf{x} \in X. \end{aligned} \tag{1.29}$$

The relationship between the optimal solution to the constraint problem and the Pareto optimality of the MOP is given by the following theorems, which are valid for both MOLP and MONLP cases:

Theorem 1.4. If $\mathbf{x}^* \in X$ is a *unique* optimal solution to the constraint problem (1.29) for some $\varepsilon_j, \ j = 1,...,k, \ \ j \ne i$, then $\mathbf{x}^*$ is a Pareto optimal solution of the MOP. If $\mathbf{x}^* \in X$ is an optimal solution to the problem (1.29) that is not unique, then it is a weak Pareto optimal solution to the MOP.

Theorem 1.5. If $\mathbf{x}^* \in X$ is a Pareto optimal solution to the MOP, then $\mathbf{x}^*$ is an optimal solution to the constraint problem (1.29) for some $\varepsilon_j, \ j = 1,...,k, \ \ j \ne i$.

The constraint method is widely used to integrate data derived from several geophysical methods. The main principle of such integration is to invert the data of one method while treating the information from the other methods as constraints. Theorem 1.5 states that only a weak Pareto optimal solution can be guaranteed if the uniqueness theorem cannot be proved for the main method selected. The solution to the constraint problem depends greatly on the values of $\varepsilon_j, \ j = 1,...,k, \ \ j \ne i$. If they are too small, the constraint problem can cease to be feasible. It is also important to note that the lower limit of resolution of the problem of joint inversion corresponds in this case to the lower limit of resolution of the problem regarded as the main objective, i.e. it is not necessarily the minimal one. These properties of the constraint method must be taken into consideration when trying to apply it to the integrated interpretation of multimethod geophysical data.

1.3.4 The weighted minimax method for the scalarization of a MOP

The second scalarization method that is not sensitive to non-convexity in the objective functions and restrictions to the MOP is the weighted minimax method (Bowman, 1976). This implies solving the following problem instead of the original MOP (1.23):

$$\begin{aligned} &\text{minimise } f(\mathbf{x}) = \max_i w_i f_i(\mathbf{x}), \quad i = 1,...,k \\ &\text{subject to } \mathbf{x} \in X \ , \end{aligned} \tag{1.30}$$

where $\mathbf{w} = (w_1,...,w_k) \geq 0$ is the vector of the weighting coefficients. The relationship of the optimal solution to the weighted minimax problem and the Pareto optimality of the MOP are given by the following two theorems, which are valid for both the MOLP and MONLP cases:

Theorem 1.6. If $\mathbf{x}^* \in X$ is a *unique* optimal solution to the weighted minimax problem (1.30) for some $\mathbf{w} = (w_1,...,w_k) \geq 0$, then $\mathbf{x}^*$ is a Pareto optimal solution to the MOP. If the solution to the weighted minimax problem is not unique, only weak Pareto optimality can be guaranteed.

Theorem 1.7. If $\mathbf{x}^* \in X$ is a Pareto optimal solution to the MOP, then $\mathbf{x}^*$ is an optimal solution to the weighted minimax problem (1.30) for some $\mathbf{w} = (w_1,...,w_k) \geq 0$.

Sakawa (1993) pointed out that the weighted minimax method solves the problem of multiobjective optimisation better than the weighted sum method or the constraints method. In spite of this, it has the disadvantage that the solution is defined by a single objective function that has its maximum value among other ones.

1.3.5 Goal programming and compromise programming methods

The idea of goal programming was originally proposed by Charnes and Cooper (1961) for a MOLP in which it is possible to specify goals or aspiration levels for the components of the vector objective function $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}),..., f_k(\mathbf{x}))$, i.e. to define the *goal vector* $\mathbf{f}^* = (f_1^*, f_2^*,..., f_k^*)$. The *goal programming problem* can then be formulated as follows:

$$\begin{aligned} &\text{minimise } d(\mathbf{f}^*, \mathbf{f}(\mathbf{x})) \\ &\text{subject to } \mathbf{x} \in X \ , \end{aligned} \tag{1.31}$$

where $d(\mathbf{f}^*, \mathbf{f}(\mathbf{x}))$ is the distance in the objective space R^k between the goal vector $\mathbf{f}^* = (f_1^*, f_2^*,..., f_k^*)$ and $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}),..., f_k(\mathbf{x}))$ according to a selected norm $l_p, 1 \leq p \leq \infty$. The aim of goal programming is to approach as close as possible a set of specified goals that may not be simultaneously attainable. Note that it is not usually possible to specify any goal vector in geophysical inverse problems *a-priori*.

A modification of the goal programming method known as compromise programming, or *the ideal point method,* was proposed by Yu (1973) and Zeleny (1973, 1982). This is

obtained from problem (1.31) if the goal vector $\mathbf{f}^* = (f_1^*, f_2^*, ..., f_k^*)$ is replaced by an ideal point or utopia vector $\mathbf{f}^{\min} = (f_1^{\min}, f_2^{\min}, ..., f_k^{\min})$, where $f_i^{\min} = \min_{\mathbf{x} \in X} f_i(\mathbf{x}), i = 1,...,k$:

$$\begin{aligned} &\text{minimise } d(\mathbf{f}^{\min}, \mathbf{f}(\mathbf{x})) = \sum_{i=1}^{k} w_i (f_i(\mathbf{x}) - f_i^{\min})^p, 1 \le p < \infty, \text{or} \\ &\text{minimise } d(\mathbf{f}^{\min}, \mathbf{f}(\mathbf{x})) = \max_i w_i (f_i(\mathbf{x}) - f_i^{\min})^p, \ p = \infty \\ &\text{subject to } \mathbf{x} \in X \ , \end{aligned} \tag{1.32}$$

where $\mathbf{w} = (w_1, ..., w_k) \ge 0$ is the vector of the weights.

It can be proved (Zeleny, 1982) that any solution to problem (1.32) for any $1 \le p < \infty$ is Pareto optimal and any *unique* solution to problem (1.32) for $p=\infty$ is Pareto optimal. The ideal point method has the advantage that it can be applied to MOLP and MONLP cases. The Pareto optimality of the solution to problem (1.32) does not depend on the convexity of the problem.

It is important to note that the ideal point is regarded as existing only in the objective space, although the corresponding solution in the parameter space may not exist. The solution to problem (1.32) can be interpreted as that which is closest to the ideal point in the objective space. The ideal point in the problem of joint inversion of multimethod geophysical data corresponds to the point in the objective space with co-ordinates equal to the maximum values for all *a-posteriori* PDFs in (1.14).

Comparison of the ideal point method with the other scalarization techniques described above demonstrates that the ideal point method can be used for joint inversion of several geophysical data sets in both a linear and a non-linear formulation. A more detailed discussion of the ideal point method with an application to the problem of magnetotelluric impedance tensor inversion is presented in Chapter 4.

In conclusion, it is necessary to point out that a generalisation of all scalarization methods for multiobjective optimisation problems is known as the *hyperplane method* (Sakawa and Yano 1990, Sakawa 1993). Since the application of this rather complicated method requires *a-priori* knowledge about the behaviour of the vector objective functions that is not available in real geophysical inverse problems, it cannot be applied directly to the solution of non-linear problems of joint inversion of multimethod geophysical data.

1.3.6 Fuzzy multiobjective optimisation

MOP (1.23) can be treated using a fuzzy set approach as proposed by Bellmann and Zadeh (1970). The basic definitions of fuzzy set theory are given in Appendix 2. Suppose that a goal vector $\mathbf{f}^* = (f_1^*, f_2^*, ..., f_k^*)$ can be defined for the problem (1.23). If it is possible to estimate the degree of achievement of the goal f_i^* at every point in the set X by a number $\mu_{C_i}(\mathbf{x}) \in [0,1]$, then each component of the vector objective function $f_i(\mathbf{x}), i = 1,...,k$ can be associated with a fuzzy set C_i defined in the parameter space and having *a membership function* $\mu_{C_i}(\mathbf{x}) \in [0,1]$. The fuzzy set C_i is then called *a fuzzy goal*. Each constraint $g_j(\mathbf{x}) \le 0, \ j = 1,...,m$ can also be associated with a fuzzy set B_j if the degree of satisfaction of the *j*th constraint at every point in the parameter space R^n can be estimated by a number $\mu_{B_j}(\mathbf{x}) \in [0,1]$. The fuzzy set B_j is called *a fuzzy constraint.*

The solution $\mathbf{x}^*$ that satisfies all the goals and constraints then belongs to the fuzzy set that is an intersection of all the fuzzy goals and constraints, with the membership function defined in accordance with *Definition A2.9*:

$$D = \left(\bigcap_{i=1}^{k} C_i \right) \cap \left(\bigcap_{j=1}^{m} B_j \right),$$
$$\mu_D(\mathbf{x}) = \min\left\{ \mu_{C_1}(\mathbf{x}),...,\mu_{C_k}(\mathbf{x}), \mu_{B_1}(\mathbf{x}),...,\mu_{B_m}(\mathbf{x}) \right\}. \tag{1.33}$$

The solution to the multiobjective problem is then a vector $\mathbf{x}^*$ that maximises the membership function of the intersection of the fuzzy goals and constraints defined by (1.33). This solution is called *a fuzzy decision* in multiobjective optimisation theory.

A number of other ways of defining a fuzzy decision also exist. Bellman and Zadeh (1970) defined *a convex fuzzy decision* as a solution to the optimisation problem

$$\begin{aligned} &\text{maximise} \sum_{i=1}^{k} \alpha_i \mu_i(f_i(\mathbf{x})),\ \sum_{i=1}^{k} \alpha_i = 1,\ \alpha_i > 0 \\ &\text{subject to } \mathbf{x} \in X. \end{aligned} \tag{1.34}$$

The other type of fuzzy decision proposed by Zimmermann (1978) is the *product fuzzy decision*:

$$\begin{aligned} &\text{maximize} \prod_{i=1}^{k} \mu_i(f_i(\mathbf{x})), \\ &\text{subject to } \mathbf{x} \in X. \end{aligned} \tag{1.35}$$

The conditions of Pareto optimality for fuzzy decisions applying to a convex MONLP are defined by the following theorems (Sakawa, 1993):

Theorem 1.8. If $\mathbf{x}^* \in X$ is a *unique* optimal solution to problem (1.33), then $\mathbf{x}^*$ is a Pareto optimal solution to the MONLP.

Theorem 1.9. If $\mathbf{x}^* \in X$ is an optimal solution to problem (1.34) with $0 < \mu_i(f_i(\mathbf{x})) < 1$ for all $i = 1,...,k$, then $\mathbf{x}^*$ is a Pareto optimal solution to the MONLP.

Theorem 1.10. If $\mathbf{x}^* \in X$ is an optimal solution to problem (1.35) with $0 < \mu_i(f_i(\mathbf{x})) < 1$ for all $i = 1,...,k$, then $\mathbf{x}^*$ is a Pareto optimal solution to the MONLP.

It should be stressed that the construction of membership functions for fuzzy goals and constraints requires subjective estimation of the degree of satisfaction of each goal and constraint prior to solution of the MOP. As demonstrated in Paper I, *a-priori* information in geophysical inverse problems can easily be presented as a system of fuzzy constraints due to its non-probabilistic nature. On the other hand, the objective functions of the geophysical inverse problem cannot be easily presented via fuzzy sets because of the probabilistic nature of the errors in the experimental data. Two types of uncertainty in geophysical inverse problems are discussed in Section 1.4.

Investigation of the behaviour of the objective function in a geophysical inverse problem usually requires a large number of forward problem calculations, which makes it difficult to formulate the problem of joint inversion of multimethod geophysical data purely as a fuzzy decision making problem. It is demonstrated in Paper I how it is possible to combine fuzzy and probabilistic approaches in one inversion scheme using the definition of Pareto optimality.

1.4 Adequate presentation of *a*-priori information in geophysical inverse problems

As already mentioned in Section 1.2, geophysical inverse problems have to deal with uncertain information that is present both in the observed data and *a-priori* knowledge about model parameters. It is for this reason that they are traditionally formulated using the mathematical apparatus of probability theory. This approach is discussed in detail in the books by Goltsman (1982) and Tarantola (1987), who proposed that the experimental data and model parameters in geophysical inverse problems should be treated as random variables specified by probability density functions. In spite of the great popularity of the probabilistic approach, it is necessary to remember that two types of uncertainty are present in geophysical inverse problems: the uncertainty attached to the observed data, which is probabilistic as it results from random observation errors, and the uncertainty attached to the *a-priori* information, which is often of a different nature, i.e. it is connected with imperfect knowledge.

Tarantola (1987) also pointed out two common intuitive interpretations of the axioms of probability that correspond to these two types of uncertainty in geophysical inverse problems. The first is purely statistical: when a "random" physical process takes place it leads to a given outcome, so that if a great number of outcomes are observed, these can be described in terms of "probability" and “random variables”. The mathematical theory of statistics is the natural tool for analysing the outputs of a random process. The second interpretation of probability is in terms of subjective degree of knowledge of the "true" value of a given physical parameter. By "subjective" we mean that it represents the knowledge of a given individual, obtained using objective methods, but that this knowledge may vary from individual to individual because each may possess different data sets. This type of uncertainty is present in *a-priori* information. Tarantola (1987) explicitly postulated that both types of uncertainty could be adequately described using probability theory. This postulate states as follows.

Let X be a discrete parameter space with a finite number of parameters. The most general way we have for describing any state of information in X is by defining a probability (in general, a measure) over X.

One of the greatest difficulties in the probabilistic formulation of inverse problems is the construction of the multidimensional PDF for the model parameters. As a rule, the use of a multidimensional PDF is limited to the case of a Gaussian PDF (both the *a-priori* PDF and the *a-posteriori* PDF in (1.10) are Gaussian and the dependence between the parameters is described via a covariance matrix):

$$p(\mathbf{m}) = \mathbf{1}/\left[(2\pi)^{k/2}\left|\mathbf{C}_M\right|^{1/2}\right]\exp\left[-1/2\,(\mathbf{m}-\mathbf{m_0})^{\mathbf{T}}\mathbf{C}_M{}^{-\mathbf{1}}(\mathbf{m}-\mathbf{m_0})\right]$$
$$p(\mathbf{d_{obs}}\mid\mathbf{m}) = \mathbf{1}/\left[(2\pi)^{L/2}\left|\mathbf{C}_D\right|^{1/2}\right]\exp\left\{-1/2\left[g(\mathbf{m})-\mathbf{d_{obs}}\right]^{\mathbf{T}}\mathbf{C}_D{}^{-\mathbf{1}}\left[g(\mathbf{m})-\mathbf{d_{obs}}\right]\right\} \quad (1.36)$$

where $\mathbf{m_0}$ is the *a-priori* model, $\mathbf{C}_M$ and $\mathbf{C}_D$ are the covariance matrices for the model and observed data, respectively, and $|\mathbf{C}_M|$ and $|\mathbf{C}_D|$ are the corresponding determinants. The model covariance matrix $\mathbf{C}_M$ describes the uncertainties in the *a-priori* knowledge about model parameters and the correlation between them, while the data covariance matrix $\mathbf{C}_D$ describes uncertainties in the data, both theoretical (modelling) and observational (Tarantola 1987). The inverse problem can then be reduced to minimisation of the objective function

$$L(\mathbf{m}) = \left[g(\mathbf{m})-\mathbf{d_{obs}}\right]^{\mathbf{T}}\mathbf{C}_D{}^{-\mathbf{1}}\left[g(\mathbf{m})-\mathbf{d_{obs}}\right]+(\mathbf{m}-\mathbf{m_0})^{\mathbf{T}}\mathbf{C}_M{}^{-\mathbf{1}}(\mathbf{m}-\mathbf{m_0})\,. \quad (1.37)$$

The function (1.37) can be minimised effectively in the case in which $g(\mathbf{m})$ is linear and the matrices $\mathbf{C}_M$ and $\mathbf{C}_D$ are diagonal. This is the main reason why the Gaussian formulation of the inverse problem became very popular among geophysicists. In real inverse problems the assumption of a Gaussian PDF for both the experimental data and the *a-priori* information is not always valid, however. One example of an inverse problem with a non-Gaussian PDF for the experimental data is the problem of magnetotelluric impedance tensor inversion considered in Chapter 4. The difficulties connected with the presentation of *a-priori* information by means of a multidimensional PDF are also analysed in Paper I, where it is proposed that possibility distributions described via fuzzy sets can be used instead of PDFs.

This proposal is based on the alternative approach to uncertain information in real problems introduced by Lofti A. Zadeh (1965), who formulated a theory in which the objects are fuzzy sets, i.e. sets with imprecise boundaries. An element of a fuzzy set can be characterised by a degree of membership, which can be not only 1 or 0, but any number in [0,1]. Zadeh's paper proposed an alternative approach to probability theory, which had for a long time been accepted as the only way to deal with uncertain information. He proposed the conception of a fuzzy logic instead of the two-valued Aristotelian logic upon which probability theory is based. Fuzzy sets provide a method for describing the uncertainty that is usually present in natural languages, for instance.

After the publication of Zadeh's paper a number of non-probabilistic theories of uncertainty were introduced by various authors in the 1970s. In particular, several theories that generalise or complement probability theory were introduced, including a probability theory of fuzzy events (Yager, 1984), a theory of Sugeno measures (Sugeno, 1977), the Dempster-Shafer theory of evidence (Shafer, 1976) and a possibility theory developed on the basis of fuzzy set theory (Zadeh, 1978).

The fundamental difference between possibility and probability was formulated by Zadeh (1978) in terms of the possibility/probability consistency principle, according to which an event should first be possible before being probable, i.e. a high degree of possibility does not imply a high degree of probability, nor does a low degree of probability imply a low degree of possibility. However, if an event is not possible, it is also improbable. An example from the paper by L. Zadeh (1978) will illustrate the difference between probability and possibility better than any formal definition. Consider

the statement: "Hans eats X eggs for breakfast", with X taking values in $U = \{1,3,4,...\}$. A possibility distribution $\pi_X(U)$ may be associated with the degree of ease with which Hans can eat *u* eggs, i.e. a possibility that he can eat 1,2,3 or 4 eggs for breakfast can be assumed to be 1. But observing Hans over a period of time, 100 days, for example, we may find that he usually eats only 1-2 eggs for breakfast. That makes it possible to associate the probability distribution $P_X(U)$ with X, which takes the values $P_X(1) = 0.5$ and $P_X(2) = 0.5$.

All theories of uncertainty are similar to the classical probability theory in the sense that they describe a type of uncertainty and use the interval [0,1] for their measures. The differences appear mainly in the way in which these measures are formulated. The concept of a fuzzy measure introduced by Sugeno (1977) and developed by Wang and Klir (1992) and Klir and Yuan (1995) provides a universal framework within which various special classes of measures can be formulated, including the classical probability measures. The brief description of the main types of fuzzy measures is given in Appendix 1.

The variety of theories of uncertainty follows from the empirical fact that uncertainty in real problems can differ in nature. Klir (1990) classified all the uncertainty types into two main groups, *fuzziness* and *ambiguity*, which can in turn be subdivided into *discord (or strife)* and *non-specificity*. Fuzziness results from a lack of definite or sharp distinctions, i.e. from imprecise boundaries of fuzzy sets. This type is treated in fuzzy set theory and is estimated by a measure of fuzziness. Ambiguity is associated with any situation in which it remains unclear which of several alternatives should be accepted as the genuine one. Discord or strife expresses conflicts among the various sets of alternatives and can be measured by classical Shannon entropy in information theory. The corresponding uncertainty measure in possibility theory is called possibilistic strife. Non-specificity (or imprecision) is connected with the sizes (cardinalities) of relevant sets of alternatives and is treated in various theories, including classical set theory, fuzzy set theory and possibility theory (see Klir and Yuan 1995 for a more detailed description of the variety of uncertainty measures addressed in these theories).

As follows from their mathematical properties, possibility, necessity, probability and other fuzzy measures do not overlap and each is suitable for modelling certain types of uncertainty. Probability theory is an ideal tool for formalising uncertainty in situations where evidence is based on the outcomes of a sufficiently long series of independent random experiments, i.e. it is capable of describe the uncertainty in experimental data. Possibility theory is very useful for formalising incomplete information expressed in terms of fuzzy propositions, and can be effectively used to present the *a-priori* information involved in geophysical inverse problems. Thus, two types of uncertainty described via different fuzzy measures with different mathematical properties are present in geophysical inverse problems, so that their unification in one inversion scheme needs special consideration.

The probability-possibility consistency principle formulated by Zadeh (1978) can be expressed mathematically as

$$Pro(A) \leq Pos(A), \tag{1.38}$$

where *Pro(A)* and *Pos(A)* denote the probability and possibility of event *A*, respectively. This consistency condition is called weak (Klir and Yuan, 1995). The strongest consistency condition would require that any event with non-zero probability must be fully possible, i.e.

$$Pro(A) > 0 \Rightarrow Pos(A) = 1 . \tag{1.39}$$

Equations (1.38) and (1.39) represent the lower and upper limits for the other probability-possibility consistency conditions that can be formulated. The probability-possibility consistency principle is the main demand for any possibility-probability transformation. In agreement with this principle, the possibility and necessity measures are regarded in the papers by Klir and Yuan (1995) and Dubois and Pride (1997) as the upper and lower bounds of ill-known probabilities.

One possible method accepted in decision-making theory for combining possibilistic and probabilistic information in one decision-making scheme is to formulate the problem as a purely probabilistic or possibilistic one using *probability-possibility transformation*. The main requirement for this was formulated by Klir (1990) as a set of three main principles for treating uncertainty: the principle of minimum uncertainty, the principle of maximum uncertainty and the principle of uncertainty invariance. These principles were derived from the fundamental fact that any type of uncertainty is uniquely connected with a given type of information deficiency. As Klir (1990) pointed out, uncertainty in any real problem is a consequence of information deficiency, which may arise for different reasons, i.e. the information may be incomplete, imprecise, contradictory, not fully reliable etc. This in turn results in different types of uncertainty in real problems, which can be formally described by different measures of information.

The *principle of minimum uncertainty* is used in general for narrowing down solutions that involve uncertainty (simplification problems), and states that the amount of information that is lost in the simplification process must be minimal. This principle is sometimes ignored in geophysical inverse problems, especially in cases where regularisation methods involving smoothing of a solution are applied. Another example of a situation in which the principle of minimum uncertainty is violated is the case when a Gaussian distribution is assumed for experimental data with a clearly non-symmetric histogram.

The principle of maximum uncertainty is important for any problem that involves *ampliative reasoning,* in which the conclusions are not entailed in the given premises (Klir and Yuan, 1995). It must be applied to any problem in which a prediction is made from a pre-defined model, including geophysical data inversion. The principle can be formulated as the following requirement: in any ampliative inference, use all the information available but make sure that no additional information is added. The principle of maximum uncertainty in classical information theory is expressed as the well-known principle of maximum entropy, or more generally, the principle of minimum cross-entropy. This approach provides the main tool for constructing unknown multidimensional PDFs from constraints on model parameters. The disadvantage of the maximum entropy/minimum cross-entropy approach arises from the fact that the Shannon entropy used to estimate uncertainty in information theory is a non-linear function. If the number of constraints is large enough, the maximisation of Shannon

entropy becomes a complicated non-linear optimisation problem. For this reason the maximum entropy approach is seldom used to estimate an unknown PDF in practical geophysical investigations.

When we approximate one model with another formalised in terms of a different mathematical theory, this basically means that we want to replace one type of information with an equal amount of information of another type: no information should be added or eliminated solely by converting from one type to another. This requirement is called the *uncertainty invariance principle* (Klir, 1990). Klir and Yuan (1995) analysed various possibility-probability transformations known in the literature from the point of view of uncertainty invariance and demonstrated that there exists a class of uncertainty invariant transformations that are unique in both directions. Unfortunately, the majority of these were developed for problems with a finite and relatively small number of alternative solutions and cannot be applied directly to geophysical inverse problems with a multidimensional parameter space in which the number of possible solutions is relatively large.

An alternative approach to combining the two types of uncertainty in one inversion scheme that satisfies all three uncertainty transformation principles described above is proposed in Paper I. This utilises the definition of Pareto optimality that was considered in the previous section. Thus the combining of various types of uncertain information in one decision-making scheme can be an additional area of application for multiobjective optimisation techniques.

2 Joint interpretation of seismic and gravity data using the relationship between rock density and seismic velocity

2.1 The density-velocity relationship as the main condition for joint interpretation of seismic and gravity data

The main condition for the joint interpretation of experimental seismic and gravity data is the established dependence between seismic velocity and rock density. This relationship, a consequence of elasticity theory, states that the velocity of seismic body wave propagation in an elastic medium depends on the elasticity tensor and the material density. As the elasticity tensor in its most common form has 81 components, 21 of which are independent, direct application of relationships from elasticity theory to the joint interpretation of seismic and gravity data is difficult.

The traditional way of performing joint interpretation of seismic and gravity data is based on the well-known correlation between rock density and seismic velocity established as a result of numerous laboratory measurements and approximated by a linear regression curve on a density-velocity diagram. The correlation between rock density and compressional wave velocity was analysed by Birch (1961) under the assumption that the latter in an isotropic medium depends primarily on the mean atomic mass and material density. In its most common form, the Birch equation connecting compressional wave velocity V_p with rock density ρ can be written as

$$V_p = a + b\rho\,, \tag{2.1}$$

where a and b are empirical constants. The main conclusion reached by Birch was proved theoretically by Anderson (1967). Since then many researchers have investigated the density-velocity relationship under laboratory conditions and calculated linear and more sophisticated non-linear regressions for it. Numerous relationships between density and V_p have been compiled from measurements made under laboratory conditions and in boreholes for various types of rocks from different geological provinces and at different pressures and temperatures. Detailed analyses of these results have been provided by

Barton (1986), Krasovsky (1981) and Schön (1998), for example. One of the most popular correlation curves used in the joint interpretation of seismic and gravity data is the Nafe-Drake curve (Ludvig et al., 1970), while more recently, Christensen and Mooney (1995) described both linear and non-linear relations linking compressional wave velocity and density in crustal and upper mantle rocks. A non-linear relationship connecting density with both compressional and shear wave velocities was obtained by Khalevin et al. (1986):

$$\rho = 2.66 - 0.107V_p - 0.0535V_s + 0.026V_pV_s + 0.0463(V_p{}^2 - 1.3333V_s{}^2)\,. \quad (2.2)$$

This large amount of information on the density-velocity relationship is usually used for the joint interpretation of gravity and seismic data in a standard manner. The area under study is divided into blocks in which density is assumed to be constant, and the density in each block is calculated from the velocity in accordance with an *a-priori* density-velocity relationship. The initial geometry of the blocks is defined from seismic data and then more precise values are obtained for the block densities and their boundaries by inverting measured gravity data, or more often by trial-and-error fitting of the calculated and observed gravity fields. Such an approach requires a detailed knowledge of the structure and velocity distributions in the region, which can be provided by seismic wide-angle reflection and refraction experiments, for example. This is why joint interpretation of gravity and seismic data with block parameterisation and a pre-defined density-velocity relationship is fairly popular in deep lithospheric investigations. It is used in the recent paper by Nielsen and Jakobsen (2000), for example, who present an algorithm for the joint inversion of wide-angle reflection and refraction travel times and gravity data that uses the density-velocity relationship of Christensen and Mooney (1995).

In spite of its popularity, the approach described above entails a number of problems. Those connected with the block parameterisation of the density model are analysed in Papers II and III. Another important problem is that the empirical density-velocity relationships for all types of lithospheric rocks demonstrate significant scattering around the mean value, so that the density-velocity relation must be regarded as a statistical dependence rather than as a functional relationship. But this statistical nature is very difficult to take into consideration in algorithms with a pre-defined density-velocity function.

The problem of adequate presentation of the density-velocity relationship for the purpose of joint interpretation of seismic and gravity data was investigated by Krasovsky (1981), who demonstrated that the linear density-velocity regression curves compiled for different types of rocks from various geological provinces and units could differ from the average regression and proposed that appropriate corrections should be introduced into the average density-velocity regression prior to interpretation. Christensen and Mooney (1995) also pointed out deviations in the density-velocity relationships compiled for different geological provinces.

2.2 Formulation of the inverse gravity problem using the density-velocity relationship

The alternative approach to joint interpretation of seismic and gravity data described in Papers II and III is an effort to treat the density-velocity relationship as a case of statistical dependence. The idea of the method was proposed in the book by Karatayev and Pashkevich (1986) and then developed in the paper by Karatayev and Kozlovskaya (1997). The method is based on the concept of a multiresponse geophysical body described in Section 1.1, i.e. the whole geological section under study is considered as a single multiresponse geophysical body *T* for which there exists an unknown dependence between seismic velocity and density that depends mainly on the PT (pressure and temperature) conditions existing within the section, the rock composition and the geology of the region concerned. If only the compressional wave velocity is known, from wide-angle reflection and refraction experiments, the equation may be

$$\rho(x,y,z) = a(x,y,z) + b(x,y,z)V_p(x,y,z)\,, \qquad (x,y,z)\subset T, \qquad (2.3)$$

where $a(x,y,z)$, $b(x,y,z)$ are unknown non-linear functions depending on pressure and temperature, rock composition and tectonic differences within the body *T*, $V_p(x,y,z)$ is compressional wave velocity and $\rho(x,y,z)$ is rock density. In other words, we assume that density depends linearly on velocity at each point in the body *T*, i.e. Birch's linear relationship is valid, but the coefficients of this dependence can vary from point to point in a manner defined by pressure and temperature, rock composition and tectonic differences.

Suppose that we have additional information about the region that was obtained independently by other methods (observed magnetic field, electromagnetic sounding data, heat flow data, depth of main reflecting interfaces etc.) and can be formulated as a finite set of continuous, bounded functions $X_i(x,y,z), i = 1,..,k$.

As the body *T* is regarded as a multiresponse one, it can be assumed that information about variations in $a(x,y,z)$ and $b(x,y,z)$ within it will also be reflected in other geophysical and geological data, i.e. that $a(x,y,z)$ and $b(x,y,z)$ can be presented as:

$$a(x,y,z) = \sum_i c_i X_i(x,y,z), \quad b(x,y,z) = \sum_j d_j X_j(x,y,z)\,, \qquad (2.4)$$

where $X_i(x,y,z)$, $X_j(x,y,z)$ are continuous, bounded functions calculated from *a-priori* information using statistical, spectral or other methods, and c_i and d_j are unknown coefficients. The main requirement for selecting of these functions is that they have to differ between the various geological units in the region. The other requirement is that they have to contain both long-period and short-period functions to represent density variations of different scale in equation (2.3). The set of functions used in Paper II was selected on the basis of previous investigations into the relation between potential field anomalies and deep lithosphere structures revealed by deep seismic sounding in the Ukrainian Shield and in the Belarus - Baltic region (Karatayev and Pashkevich, 1986). The functions demonstrated that the regional gravity and magnetic field patterns and the

depths of the main seismic boundaries that could be recognized in the deep seismic sounding data do in fact different between the geological units in the region. These boundaries are usually the upper boundary of the crystalline basement, the boundary between the upper and lower crust and the Moho boundary, the values of which, together with the regional magnetic field and the derivatives of all of these, are used in Paper II to construct the long-period and short-period functions $X_i(x, y, z)$, $X_j(x, y, z)$ in (2.4).

When both the compressional and shear wave velocities are known from the seismic data, the density-velocity relation can be composed based on eq. (2.2), as proposed in Paper III, i.e. it can be presented as a linear combination of non-linear functions of both V_p and V_s complemented with other data.

Substituting (2.4) into (2.3) and performing all the necessary algebraic operations, we obtain a basic approximation for the density distribution within the body T:

$$\rho(x, y, z) = \sum_k \{A_k U_k(x, y, z)\}, \tag{2.5}$$

where $U_k(x,y,z)$ are continuous, bounded functions, calculated from *a-priori* information represented by functions $X_i(x, y, z)$ and the seismic wave velocity, and A_k are unknown coefficients. Equation (2.5) can then be substituted into the forward problem operator for the gravity field. To calculate the gravity effect caused by the body T in the observation area $Q=(x_0,y_0,z_0)$, we have:

$$V_z(x_0, y_0, z_0) = G\int_T \frac{\rho(x, y, z)(z - z_0)}{r^3} dT, \tag{2.6}$$

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2},$$

where G is the gravitational constant and $dT = dxdydz$.

After substituting (2.5) into (2.6), and after all the necessary simplification and numerical integration, we obtain a simple linear expression for the gravity effect $V_z(x,y)$ caused by the density distribution (2.5) within the body T:

$$V_z(x,y) = \sum_k \{A_k W_k(x,y)\},\ k = 0,..., K, \tag{2.7}$$

where $W_k(x, y) = \sum\sum\sum \{L_{ijm}(x,y) U_k(i\Delta x, j\Delta y, m\Delta z)\}$ and the coefficients A_k are unknown. The function $L_{ijm}(x,y)$ denotes the gravity effect of a rectangular 3-d prism with unit density, $\rho(x,y,z)=1$. The centre of the prism is located at a node on a regularly spaced grid defined in the area T to calculate the integral (2.6) numerically. The symbols Δx, Δy and Δz denote steps on the grid along the x, y and z-axes, respectively. In this case the coordinates of the prism centre are $(i\Delta x, j\Delta y, m\Delta z)$ and the sizes of the prism along the x, y and z-axes are Δx, Δy and Δz, respectively. The values Δx, Δy and Δz have to be small enough to ensure presentation of the velocity distribution in the equation (2.3) with the same spatial resolution as was obtained for the seismic data interpretation and the necessary accuracy of the gravity effect calculation.

The functions $W_k(x,y)$ in (2.7) need to be calculated only once before the inverse problem is solved, i.e. the forward gravity problem calculation is performed only once.

The edge effects in the calculated theoretical gravity effect are eliminated by continuation of the area T to infinity. A positive offset in the calculated gravity effect due to the use of absolute values for the density in (2.6) is corrected by adding an unknown constant A_0 to equation (2.7). The corresponding function to be included in equation (2.7) is then $W_0(x, y) \equiv 1$. This correction results in one additional unknown parameter in the inverse problem.

The parameter vector of the inverse gravity problem **m** consists of the set of coefficients A_k. This can be obtained by minimising the difference between the gravity field $\Delta g(x,y)$ observed in the area Q and the calculated model gravity field $V_z(x,y)$ in the l_2 norm, i.e. under the assumption of a Gaussian PDF for the error in the observed data:

$$\left\| \Delta g(x, y) - V_z(x, y) \right\|_{l_2} \rightarrow \min_{A_k} . \tag{2.8}$$

The proposed approach makes it possible to formulate the inverse gravity problem as a linear one with a fixed and relatively small number of model parameters:

$$\mathbf{Gm} = \mathbf{d} , \tag{2.9}$$

where $\mathbf{m} = (A_0, A_1, ..., A_k)$, $k = 0, ..., K$, $\mathbf{G}$ is the matrix composed of the values of functions $W_k(x,y)$ calculated at the points of observation of the gravity data, and $\mathbf{d} = (d_1, d_2, ..., d_N)$ is the vector of the observed gravity data.

An equation approximating the unknown density-velocity relationship is then obtained as a solution to the linear least squares problem and used to calculate the unknown density distribution (2.5). The model field (2.7) is the effect of the gravity sources for which the density-velocity relationship (2.5) is satisfied.

It should be noted that an approximation of the density models with predefined orthogonal functions (Legendre polynomials, trigonometric functions etc.), aiming at linearisation of the inverse gravity problem, has also been used by other authors, among them Bott (1973) and Zidarov (1990), who have demonstrated that these functions correspond to density distributions producing the same gravity effect as the true one but are often geologically meaningless. Approximation of the density distribution by sinusoidal signals, or other orthogonal functions can be a useful tool for formal analysis of the gravity field, but the possibilities for representing real density distributions in this way are limited. It is known that any signal, including a seismic boundary, can be expanded into a Fourier series, but one needs a large number of terms to represent a signal of arbitrary form by trigonometric functions. Thus real geophysical information, i.e. velocity distribution, seismic boundaries and additional geophysical fields, are more useful for the approximation of density distributions, i.e. they simply contain the information in a more compact form. A similar approach is used in factor analysis, for example, where the experimental data are expressed via a linear combination of known factors that are usually selected from real data sets.

It can be concluded from the above that the main difference between the proposed procedure of graviseismic modelling and the traditional approach with an *a-priori* defined density-velocity relationship is that the density-velocity relationship is used for parameterisation of the unknown density model and then obtained as a solution to the inverse gravity problem. An important advantage of such parameterisation is that the

number of model parameters remains the same for 3-D and 2-D models. The method does not require *a-priori* knowledge of any density-velocity relationship and can be used to interpret data from different geological provinces.

2.3 Treatment of non-uniqueness of the inverse gravity problem

Although a least square solution to the inverse gravity problem (2.9) always exists and is unique for a given set of functions $U_k(x,y,z)$, the fundamental non-uniqueness of the problem (Skeels, 1947) cannot be avoided, and this can result in a geologically meaningless solution. The different subsets of functions in (2.8) can correspond to entirely different density distributions that produce very similar effects in the observation area. Errors in the observed data also contribute to the non-uniqueness of the problem. As demonstrated in Chapter 1, the non-uniqueness of the inverse problem can be reduced if the set of feasible solutions is constrained by *a-priori* information. This information is introduced into the present algorithm in a manner that can be called "interactive". The problem (2.8) is solved with different subsets of functions $U_k(x,y,z)$ giving different equations (2.5) and different density sections. The final solution is selected on the basis of a visual analysis of the density section in accordance with several qualitative criteria representing *a-priori* knowledge about the unknown density distribution:

a) The density section must be meaningful. Examples of meaningless solutions are density distributions that look like a chessboard or alternation of vertical lines, or a density distribution in which density decreases systematically with depth.
b) It must not differ very much from the initial velocity distribution, i.e. it must preserve the main features revealed by the seismic data modelling.
c) The density values in the upper part of the section must be in agreement with the *a-priori* petrophysical data on the density of the uppermost crystalline basement and sedimentary cover.
d) The density values for the upper mantle below the Moho boundary must be in agreement with petrophysical data regarding the density of the upper mantle rocks.

The *a-priori* conditions c) and d) can in principle be formulated as constraints on density values, but their direct inclusion in the inverse problem (2.8) requires their re-formulation with respect to the coefficients A_k, which is difficult to do. That is the reason why they are treated as qualitative in the present algorithm. It would also be wrong to think that the condition c) requires the density section to be completely identical to the velocity section. It is known that data from wide-angle reflection and refraction experiments allow horizontal and sub-horizontal interfaces to be revealed but usually contain very little information about vertical or sub-vertical structures. Gravity data, on the other hand, are sensitive specifically to vertical structures, i.e. the calculated density section can contain information that is not present in the initial velocity section. Thus only qualitative analysis of the difference between the initial velocity section and the resulting density section is used in the present algorithm. In spite of some additional calculations that are necessary for proper qualitative analysis of possible solutions to problem (2.8), the approach allows a solution to be obtained that agrees not only with the seismic data but

also with the *a-priori* petrophysical information and our common understanding of the density distribution in the lithosphere.

2.4 Application of the technique to the joint interpretation of gravity and wide-angle reflection and refraction data

The approach and its application to joint interpretation of the seismic and gravity data obtained for the EUROBRIDGE'95 profile in Lithuania and the SVEKA profile in Finland are described in Papers II and III. The method has also been used to interpret the LT-7 and TTZ profiles in Poland (Kozlovskaya and Yliniemi, 1999) and EUROBRIDGE'96 in Belarus (Kozlovskaya et al., 1999). The seismic models developed by J. Yliniemi were used in the procedure of joint interpretation. The results of testing the algorithm with real data allow us to conclude that joint interpretation of seismic and gravity data by means of the above technique makes it possible to compile the density section with the same spatial resolution as can be achieved in velocity models obtained from wide-angle reflection and refraction data. The final density model selected from the set of possible solutions inherits the main features of the initial one. The joint interpretation of seismic model and gravity data also makes it possible to distinguish vertical and sub-vertical structures and discontinuities that cannot be recognised in seismic data alone.

The density-velocity diagrams for the crust and upper mantle that were obtained from the density-velocity model of the EUROBRIDGE'95 and SVEKA profiles described in Papers II and III are presented in Fig. 2.1(a-f). It is seen that the calculated density values are scattered over the corresponding velocity values, i.e. the algorithm allows a non-unique density-velocity relation to be modelled. The density-velocity diagrams obtained from the joint interpretation of seismic and gravity data are similar to those compiled from the laboratory measurements. The breaks in the density-velocity diagrams in Fig 2.1 are due to velocity jumps at seismic boundaries in the initial seismic models. Comparison of diagrams for the EUROBRIDGE'95 and SVEKA profiles demonstrates that scattering of the density-velocity curve is much greater when both *Vp* and *Vs* are used to model density.

A resolution analysis can be performed on the minimum least square problem (2.9) in the standard way (Menke, 1989), i.e. a covariance matrix can be calculated for the parameter vector estimates directly from the matrix $\mathbf{G}$ and the variance of the observed data vector:

$$[\mathrm{cov}\,\mathbf{m}] = \sigma_d^2 \left[\mathbf{G}^{\mathbf{T}}\mathbf{G}\right]^{-1}, \tag{2.10}$$

where $[\mathrm{cov}\,\mathbf{m}]$ is the covariance matrix for the vector $\mathbf{m}$ and σ_d^2 is the variance of the observed data under the assumptions that the components of the data vector are uncorrelated and all of equal variance and that the matrix $\mathbf{G}$ contains no error. The covariance matrices of the parameter estimates for the EUROBRIDGE'95 profile and the SVEKA profile are given in Tables 1 and 2, respectively.

Table 1. Covariance matrix of parameter estimates for the EUROBRIDGE'95 profile

	A0	A1	A2	A7	A12	A17	Parameter estimates
A0	0.002013	9.28E-05	-0.0003	-1.25E-05	6.76E-07	-1.3825	-0.679
A1	9.28E-05	4.84E-05	-5.99E-05	1.98E-07	4.83E-07	-0.0170	0.239
A2	-0.0003	-5.99E-05	9.37E-05	9.84E-07	-1.01E-06	0.13556	-0.164
A7	-1.25E-05	1.98E-07	9.84E-07	1.13E-07	9.23E-09	0.0107	0.002
A12	6.76E-07	4.83E-07	-1.01E-06	9.23E-09	2.87E-07	0.0086	0.007
A17	-1.3825	-0.0170	0.1355	0.01076	0.0086	1416.8	779.37

Table 2. Covariance matrix of parameter estimates for the SVEKA profile

	A0	A1	A11	A12	A13	A14	A16	A17	Parameter estimates
A0	5.0401	0.3384	0.0789	1.1325	-0.2813	-1.2083	-3.7284	-20.168	-1.087
A1	0.3384	0.0466	0.0059	0.1789	-0.0497	-0.1716	-0.3039	-0.4927	0.127
A11	0.0789	0.0059	0.0016	0.0200	-0.0051	-0.0210	-0.0603	-0.8490	0.134
A12	1.1325	0.1789	0.0200	0.7603	-0.2141	-0.7156	-1.0658	0.6284	-0.011
A13	-0.2813	-0.0497	-0.0051	-0.2141	0.0607	0.1999	0.2764	-0.3964	0.0034
A14	-1.2083	-0.1716	-0.0210	-0.7156	0.1999	0.6797	1.0946	0.4217	0.012
A16	-3.7284	-0.3039	-0.0603	-1.0658	0.2764	1.0946	2.8786	13.739	-0.039
A17	-20.168	-0.4927	-0.8490	0.6284	-0.3964	0.4217	13.739	1307.7	966.73

As matrix **G** in problem (2.9) is calculated from the *a-priori* data on the velocity distribution and seismic boundaries, a high quality initial velocity model is a necessary condition for successful gravity data interpretation. The full waveform inversion of wide-angle reflection and refraction data that is capable of providing high velocity resolution is still not used very widely, as it results in a very large-scale problem, and for this reason the seismic models for the profiles considered here were obtained by the trial-and-error fitting of observed and calculated travel times complemented with comparisons of the amplitudes of synthetic and experimental seismograms. This additional seismogram analysis increases the velocity resolution of models compiled from wide-angle reflection and refraction data, and as a rule gives better results than inversion of travel times only (Morozova and Pavlenkova, 1995). The velocity uncertainty in parts of profiles with good ray coverage is estimated to be 0.1 km/s. The spatial vertical resolution of the velocity model and the resulting density model in this case is estimated to be 1 km for the uppermost crust and 2 km for the lower crust and upper mantle (EUROBRIDGE Seismic Working Group, 1999). The size of the lateral inhomogeneities that can be resolved by such an interpretation depends on the geometry of the experiment and is approximately 10-30 km in the parts of the model with good ray coverage. As the velocity in the upper mantle and at the ends of the profiles is usually poorly recovered by wide-angle reflection and refraction data, the corresponding density values are less reliable than in the central part.

One important property of the proposed technique is that it allows bodies to be detected within the section under study for which the common density-velocity relationship is violated. The detection of such zones is based on analysis of the residual gravity field $\Delta\gamma(x) = \Delta g(x) - V_z(x)$ in the manner described in Section 1.1.

The integrated density-velocity model along the EUROBRIDGE'95 profile presented in Paper II is an example of a multiresponse body, because good fitting of the observed and model gravity field was achieved over the whole observation area, while the velocity-density model along the SVEKA profile described in Paper III is an example of the combination of a multiresponse body for which the common relationship between density, *Vp* and *Vs* holds with two bodies for which this relationship is violated. The latter correspond to two positive anomalies in the residual field that coincide spatially with the Kuhmo Greenstone belt and Svecofennian schists in the vicinity of the Lake Ladoga-Bothnian Bay Zone. In spite of the fact that petrophysical examinations of rocks from these areas demonstrate that the density at the surface is rather high (Elo et al., 1978), the velocity sections along the SVEKA profile do not indicate high velocity corresponding to these geological structures. It can be concluded that these are the places where the common density-velocity relationship is violated, i.e. they can be regarded as single-response geophysical bodies. It is shown in Paper III how the residual $\Delta\gamma(x)$ can be interpreted separately, aiming at finding parameters for single-response bodies.

Another example of violation of the density-velocity relationship was revealed in the EUROBRIDGE'96 profile, extending across one of the most important lithospheric discontinuities in the East European Craton, the Fennoscandia-Sarmatia junction zone. Interpretation of the EUROBRIDGE'96 seismic refraction and wide-angle reflection profile revealed a high P-wave velocity zone at a shallow depth beneath the southeastern part of the Central Belarus belt (SECB) which was explained by presence of mafic rocks (EUROBRIDGE Seismic Working Group, 1999). This explanation apparently contradicts the gravity data observations, which do not indicate any pronounced positive Bouguer anomaly in this place. One possible explanation of this disagreement can be that the high velocity beneath the SECB is caused by seismic anisotropy. This hypothesis explains not only the high velocity, but also the absence of a positive gravity anomaly (Kozlovskaya et al., 1999).

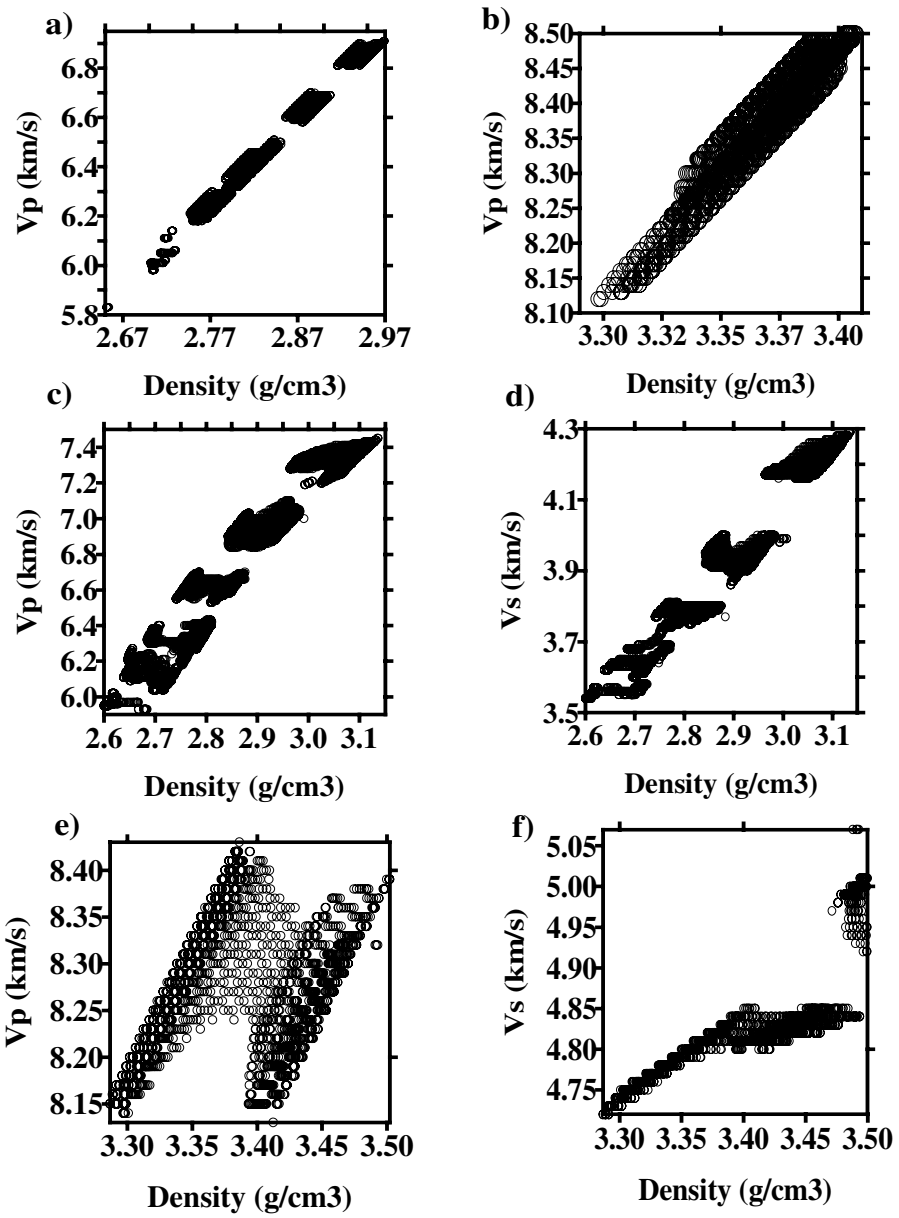

Fig. 1. Density-velocity diagrams: a) Density-Vp diagram for the crust obtained for the EUROBRIDGE'95 profile; b) Density-Vp diagram for the upper mantle obtained for the EUROBRIDGE'95 profile; c) Density-Vp diagram for the crust obtained for the SVEKA profile; d) Density-Vs diagram for the crust obtained for the SVEKA profile; e) Density-Vp diagram for the mantle obtained for the SVEKA profile; e) Density-Vs diagram for the mantle obtained for the SVEKA profile.

2.5 Seismic anisotropy as the reason for violation of the density-velocity relationship

The proposed explanation for the violation of the common density-velocity relationship agrees with the results of recent investigations into the physical properties of crustal and uppermost mantle rocks, which have demonstrated that all of these are more or less seismically anisotropic (Babushka and Cara, 1991). For highly anisotropic rocks, the density-velocity relationship is violated due to the dependence of seismic velocity on the direction of propagation.

In their recent investigation, Meissner and Rabbel (1999) pointed out that any geological ordering processes resulting in structural ordering of rocks and layering, such as sedimentation or ductile flow, should also cause seismic anisotropy. This conclusion is supported by the results of recent seismic and petrophysical investigations aimed at studying the continental lower crust. Regional-scale sub-horizontal layering accompanied by seismic anisotropy in the lower continental crust has been revealed by seismic surveys in many places (Rabbel and Mooney, 1996, Rabbel et al., 1998). Investigations into the exposed laminated lower crust in the Ivrea zone (Burke and Fountain, 1990) and beneath Urach (Rabbel et al., 1998), supported by local xenolith studies, also demonstrated that the laminated reflecting pattern in the lower crust can be accompanied by seismic anisotropy caused mainly by the preferred orientation of anisotropic minerals. Thus, information on seismic anisotropy can be used as an additional constraint in the geological interpretation of geophysical models.

The existence of similar regional-scale reflecting and anisotropic structures in the brittle middle and upper crusts of old, cold regions is not so obvious. Meissner and Rabbel (1999) pointed out that any previous lower crustal lamellae transferred into the brittle regime might have been destroyed and deformed by tectonic stresses. This conclusion is partly supported by experimental results from the German Continental Deep Drilling programme (KTB), and seismic surveys in the same area have demonstrated numerous local variations in seismic anisotropy caused by preferred mineral orientation (Rabbel, 1994), but have not revealed any uniform regional pattern of anisotropy.

One known example of regional-scale anisotropic structures in the brittle upper crust concerns the mylonitic shear zones that are often associated with high reflectivity (Jones and Nur, 1984, Kern and Wenk, 1990). Mylonitic zones are often several kilometres thick in surface exposures and consist of multiple undulating bands of variable fabric, composition and intensity of deformation. Jones and Nur (1984), investigating the mylonitic rocks of the Wind River thrust in Wyoming, found that the rock fabrics are mainly parallel or near parallel to the regional trend of the fault zone. This example demonstrates that seismic anisotropy on a rather large scale can also exist in the brittle crust, but it is not formed everywhere and it is associated with zones of tectonic contacts where structural ordering of rocks takes place. Meissner and Rabbel (1999) pointed out that the main condition for the long-term survival of such structures is that no new thermal event should have taken place since their formation.

The results of seismic refraction and wide-angle reflection experiments are usually interpreted in terms of isotropic velocity models in which velocity variations are

explained only by differences in rock composition. The possible influence of seismic anisotropy on seismic velocity is usually ignored. As a result, a number of the seismic models interpreted contain details that are in contradiction with the gravity data. Such unexplained details are very often located in the zones of important tectonic boundaries, i.e. in places, which are targets of seismic experiments. One such unexplained phenomenon is the famous “Erbendorf body” in the vicinity of the KTB drilling site. This is a zone of high reflectivity accompanied by a thin layer of high velocity in the middle crust detected in the central part of the DEKORP 4 profile (Gebrande et al., 1989). The authors pointed out that such high velocity in the middle crust could have a density contrast of about 0.3 g/cm^3 with the surrounding crust, which could result in an additional gravity anomaly of 14 mgal. It was proposed that the laminar structure with its smaller average density contrast may explain both the high velocity and the reflectivity of the “Erbendorf body”. Such a laminated structure can also result in weak transverse anisotropy.

A second similar example was revealed by geophysical investigations within the Saxo-Thuringian belt and is connected with the unexplained origin of the Saxonian high-pressure granulites (DEKORP and Orogenic Processes Working Groups, 1999). Seismic reflection and refraction profiling revealed a broad, dome-like reflective zone in the middle crust that simultaneously had very high P-wave velocities (up to 7.0 km/s). The zone can be seen in several near-parallel seismic profiles, i.e. it is laterally persistent over a broad area. Although high P-wave velocities were detected, no corresponding gravity maximum was observed. As in the previous example, the origin of this phenomenon cannot be explained in terms of isotropic velocity.

The above examples demonstrate that zones with exposed disagreement between velocity and density in the lithosphere, where the common density-velocity relationship is violated, can provide important information for the correct geological interpretation of seismic models. The ability to reveal these zones can be considered an important advantage of the proposed technique for joint interpretation of seismic and gravity data.

3 Relationship between elastic and electrical properties of rocks as a base of joint interpretation of seismic and EM data

3.1 Problem formulation

Joint quantitative interpretation of electromagnetic and seismic experimental data sets can be performed only if we can suppose that they contain information about geological objects for which the dependence between electrical and elastic properties exists. That is why the first step to joint inversion of seismic and EM data must be establishing the common factors that affect both elastic and electrical rock properties and determination of the situations in which joint quantitative interpretation of seismic and electromagnetic data can be useful.

The connection between elastic and electrical rock properties is not so obvious, as in the case of density-velocity relationship, although the results of some geophysical experiments demonstrate that there exists a correlation between high conductivity and low velocity zones in the porous rocks saturated with water and in the partially molten rocks in the mantle. Thus, the solid-liquid rock systems can be a class of objects for which the quantitative relationship between electrical and elastic properties can be established and joint inversion of seismic and EM data can give better results than their separate inversion. In the present work the problem of relationship between elastic and electrical rock properties was considered from two different points of view:

a) Using theoretical modelling of elastic and electrical properties of solid-liquid rock systems.
b) Using analysis of petrophysical data of the German Continental Deep Drilling Program (KTB).

3.2 Factors influencing both elastic and electrical properties of lithospheric rocks revealed by previous theoretical modelling

3.2.1 Theoretical modelling of elastic properties of solid-liquid rock mixtures

It is known that fractures, cracks and other defects of the rock composing minerals change the elastic properties of rocks, i.e. the elastic wave velocity decreases due to defects. The other effect is dependence of elastic properties on pressure: seismic velocity increases with depth in the uppermost crust as cracks and fractures are closed due to lithostatic pressure. There exist a great amount of theoretical rock models describing the elastic properties of saturated porous rocks via their dependence on porosity that were developed mainly for sedimentary rocks.

The simplest velocity-porosity relationship can be obtained from the fact that velocity is related to density. The former, in turn, is linearly related to porosity by the simple relationship: $\rho = \rho_1\phi + \rho_2\phi_2$ where the subscripts 1 and 2 refer to fluid and rock matrix, respectively, ρ is the wet bulk density, ϕ is porosity and $\phi_2 = 1 - \phi$.

One of the simplest theoretical-empirical velocity-porosity relationships of such a kind was obtained by Nafe and Drake (1963) for marine sediments:

$$V_b = \phi(V_1^2[1 + \rho_1 / \rho_2]\phi_2) + (\rho_2 / \rho_b)\varphi_2^N V_2^2 \tag{3.1}$$

where subscripts 1,2 and b represent property of pore fluid, dry solid material and bulk (combined) material, respectively. Setting *N* equal to 4 and 6 generates two functions that act as upper and lower bounds for most of the observed cases.

The velocity-porosity dependence for magmatic and metamorphic rocks is more complicated and cannot be described by eq. (3.1) because many of the parameters in rock physics models are strongly dependent on the details of rock microstructure.

A more precise relationship describing the influence of pore fluid on elastic properties can be found in the classical theory by Gassmann (1951) that describes elastic wave propagation in saturated porous rocks under the assumption that any relative motion between the solid rock matrix and the fluid in pores is negligible. Gassmann assumed that the pore fluid does not change the properties of the solid skeleton, i.e. shear modulus of saturated rock is equal to shear modulus of dry rock and the rock skeleton bulk modulus is the same for the dry and saturated case. The theory provides a simple expression relating the saturated rock elastic moduli to the dry rock elastic moduli:

$$K_{sat} = K_{dry} + \frac{\left(1 - \frac{K_{dry}}{K_M}\right)^2}{\frac{\phi}{K_p} + \frac{1-\phi}{K_M} - \frac{K_{dry}}{K_M^2}}, \qquad (3.2)$$

$$\mu_{sat} = \mu_{dry}$$

where K_{dry}, K_M and K_p are the bulk moduli of dry rock, the mineral grains, and the pore fluid, respectively, and φ is the porosity. K_{sat} is the predicted bulk modulus of the saturated rock.

Then the correspondent relationship for the plane wave modulus defined by White (1983) as $M = K + 4/3\mu$ is:

$$M_{sat} = M_{dry} + \frac{\left(1 - \frac{K_{dry}}{K_M}\right)^2}{\frac{\phi}{K_p} + \frac{1-\phi}{K_M} - \frac{K_{dry}}{K_M^2}} \qquad (3.3)$$

P- and S-wave velocities for fluid saturated rock are then:

$$V_p = \sqrt{\frac{M}{\rho}} \quad , V_s = \sqrt{\frac{\mu}{\rho}} \qquad (3.4)$$

Equations (3.2-3.4) demonstrate the non-linear dependence of elastic wave velocity on porosity.

Gassmann's theory of elastic wave propagation in a fluid saturated solid was later developed by Biot (1956a,b). He included dynamic effects through connected pores in his theory and, additionally, introduced fluid viscosity and hydraulic skeleton permeability into his model. The viscous properties of the model result in attenuation and a frequency dependence of seismic wave velocities. Biot gave the relationship for a characteristic frequency f_c that divides the elastic behaviour of rock into two frequency ranges relating to the dominating effect of fluid motion:

$$f_c = \frac{\eta_f \phi}{2\pi \rho_f \ k_{hydr}} \qquad (3.5)$$

where η_f is the viscosity of the pore fluid, φ is the porosity, ρ_f is the density of the pore fluid, and k_{hydr} is the hydraulic permeability. In the “low frequency range” i.e. if $f \prec\prec f_c$ the

motion in the rock-liquid system is controlled by viscous drag against the solid. In the "high frequency range", i.e. $f >> f_c$, the inertia of fluid dominates.

Due to its relative simplicity the Gassmann-Biot theory and its later developments are widely used for theoretical calculation of the dependence between porosity and seismic velocities. As Schön (1998) pointed out, the fundamental problem in the Gassmann-Biot theory is the unknown value of the skeleton bulk modulus *K*. Calculation of velocity requires determination of the skeleton bulk modulus either experimentally or theoretically, that, in turn requires knowledge of rock skeleton microgeometry.

The other important class of solid-liquid rock models are inclusion models in which a homogeneous solid matrix containing isolated pores or cracks is modelled as an elastic solid containing inclusions of various shape and content. One important class of such models, namely, the model with ellipsoidal inclusions was initially proposed by Eshelby (1957) and later developed by Walsh (1965) for "penny shaped" cracks and by Mavko and Nur (1978) for non-elliptical cracks. The advantage of these models is that they allow to describe the seismic anisotropy in real rocks due to aligned pores and fractures. The so-called "self-consistent method" for calculation of elastic properties of a material with inclusions was developed by Budyansky and O'Connel (1976). This self-consistent method takes into consideration the effects caused by fluid flow mechanisms on the grain scale. Toksöz et al. (1976) proposed a method to calculate the average elastic and anelastic properties of a model with isolated spherical or spheroidal inclusions by studying plane wave scattering by inclusions.

There exist also a class of theoretical rock models based on packing of spheres that suit well to describe properties of unconsolidated rocks. The detailed analysis of sphere packing models can be found in Schön (1998).

The application of inclusion models and packing spheres models was always limited by their ability to model the properties of grain-to-grain contact in real rocks that can affects significantly the elastic properties. The effect of grain form and morphology of the liquid phase on elastic properties is especially strong in partially molten rocks. As Mavko (1980) pointed out, the estimate of melt fraction in partially molten rocks depends on the model used to approximate the melt geometry. Schmeling (1985) demonstrated a strong influence not only of the melt fraction, but also of the melt geometry and degree of interconnection on seismic velocity and pointed out difficulties in the application of the ellipsoidal inclusion model to approximation of real melt geometries. He proposed a model with inclusions in the form of melt tubes along grain boundaries that can better approximate the real melt geometries than models with ellipsoidal inclusions. It is necessary to note that the model did not allow to take into consideration seismic anisotropy due to preferred orientation of melt inclusions.

The effect of the model used to approximate the melt geometry on the estimated seismic velocity was demonstrated also by Faul et al. (1994) who used experimentally determined geometry of melt inclusions to calculate seismic velocities and found out that the estimates for penny-shaped melt inclusions and tubulous inclusions are different. Recent investigations of microstructure of real partially molten rocks by Mainprice (1997) and Lamoureux et al. (1999) demonstrated once again the effect of shape and orientation of melt inclusions on seismic anisotropy and one more time pointed out the difficulties connected with using traditional theoretical rock models for an approximation of real melt geometries.

The models with variable internal structure can be considered as possible candidates to model the realistic geometries of solid-liquid rock systems. Models of this class were developed in the papers by Schön (1983) and Spangenberg (1998). The latter model is based on the conception of fractal structure of grain-to-grain contact and can model such properties of solid-liquid rock system as non-ideal shape of grains and inclusions and elongated grains and pores that can result in the seismic anisotropy. The model was used in Paper 4 to calculate elastic properties of weakly porous crystalline rocks and partially molten rocks.

3.2.2 Theoretical modelling of electrical conductivity of solid-liquid rock mixtures

For porous or fractured water-bearing rocks, the electrolytic conductivity of the water itself and the interaction between solid and fluid components cause increment of conductivity. Therefore, conductivity differs for dry and saturated rocks of the same mineral composition.

The electrical conductivity of a water saturated rock is controlled mainly by the properties of the water that is an electrolyte. The conductivity of the solid matrix is negligible in most cases except in the presence of ores, graphite or clay. The conductivity of water that is present in minerals depends on its saturation with salts and varies from 5×10^{-6} (pure) to roughly 10^{2} (saturated) S/m. Mixtures of these materials can have a range of resistivities from 10^{-8} to 10^{17} Ohm m.

The situation in the partially molten rocks is not so simple because the rock bulk conductivity under high pressure and temperature depends also on the conductivity of the solid matrix, i.e. olivine conductivity. As it was demonstrated by the measurements of electric properties of polycrystalline olivine aggregates under high pressure and temperature (Roberts and Tyburczy, 1991), the geometry of grain-to grain contact also affects the bulk electrical conductivity in such aggregates.

The simplest way to describe conductivity of solid-liquid rock system was proposed by G.E. Archie (Archie, 1942). He suggested that the conductivity of brine saturated rocks is proportional to brine conductivity and, also, the specific resistivity of the water saturated rock ρ_0 is proportional to the specific resistivity of the brine ρ_w, i.e. that the brine conduction is only one conductive mechanism:

$$\rho_0 = F\,\rho_w \tag{3.6}$$

where F is formation resistivity factor. It expresses the resistivity magnification related to the brine as a result of the presence of a non-conductive matrix (formation). In other words, it shows the correlation with connected porosity. The dependence between resistivity and porosity Φ then is expressed by the first Archie's equation:

$$\rho_0/\rho_w = F = 1/\Phi^m \tag{3.7}$$

where m is an empirical quantity that can be different for various rocks. Archie noted that m has a value about 1.3 for unconsolidated sands and a range between 1.8 and 2.0 for many consolidated sandstones and called it “cementation exponent”. Archie’s equation gave the first practical relationship between a measurable property (resistivity) and an important reservoir property (porosity). Further studies showed that a better fit with experimental data has an expression in which

$$\rho_0 / \rho_w = F = a / \Phi^m \tag{3.8}$$

where a is a second empirical parameter depending on rock type. It is important to note that both empirical parameters a and m are controlled by the pore channel microgeometry. If the conducting pore channels are deformed by pressure, the rock conductivity decreases (resistivity increases), and the formation factor increases. The non-linear increase of F with increasing pressure results from the non-linear stress-strain behaviour of the rocks related to the reduction of the pore sizes. To a first approximation $F \sim p^g$ where g is called a pressure exponent. For some types of sedimentary rocks the following equation was found empirically (Schön, 1998):

$$F_p / F_{400} = a + b \log p \tag{3.9}$$

where p is the actual pressure, F_p is actual formation factor at a pressure p, F_{400} is formation factor at a pressure 400 MPa, a and b are empirical constants.

Archie’s law was primarily obtained for reservoir rocks, but it has been also applied with success to fractured igneous rocks with very low porosity and to highly porous unconsolidated rocks. The modified Archie’s law was used by Hermance (1979) and Watanabe and Kurita (1993) to estimate the effective electrical conductivity of partially molten rocks. But as a whole, the dependence of electrical resistivity of igneous and metamorphic rocks upon porosity and pressure is more complicated. It is important to remember that application of Archie's law is limited to the case when rocks can be assumed as electrically isotropic and cannot be applied to model rocks with structural electrical anisotropy.

Hashin and Shtrikman (1963) proposed another approach to estimate effective conductivity of a solid-liquid rock system in which the liquid phase is assumed to be present in idealized geometries. Each of the two phases is assumed to be homogeneous and isotropic and possible conduction effects along grain surfaces are not taken into consideration. The corresponding relationships give upper and lower boundaries of the effective conductivity of a two-phase system and are called Hashin-Shtrikman-bounds (H-S bounds):

$$\sigma_{HS^-} = \sigma_0 + \beta \left(\frac{1}{\sigma_f - \sigma_0} + \frac{1-\beta}{3\sigma_0} \right)^{-1} < \sigma < \sigma_f + (1-\beta) \left(\frac{1}{\sigma_0 - \sigma_f} + \frac{\beta}{3\sigma_f} \right)^{-1} = \sigma_{HS^+} \tag{3.10}$$

where β is volumetric fraction of fluid and σ_0, σ_f are the specific conductivities of the solid and the fluid, respectively.

There exists also a number of theoretical conductivity models developed namely for the case of partially molten rocks in which the geometry of melt films is taken into consideration (Grant and West, 1965, Waff, 1974).

A class of statistical models in which the variable degree of interconnection of the liquid phase is modelled by a general resistor network was proposed by Rink and Schopper (1968), Madden (1976) and some other authors. The resistor network approach was also used by Schmeling (1986) in his fundamental paper concerning calculation of electrical properties of partially molten rocks. He combined models with different melt geometries with resistor models and obtained correspondent dependencies of bulk resistivity on melt fraction for different pore geometries. The advantage of the model proposed by Schmeling is that both elastic and electrical properties were calculated for the same pore geometry (Schmeling, 1985,1986). The case of anisotropic rock structures was not investigated, i.e. the calculations were limited only to the case of isotropic melt geometries.

The fractal random resistor networks were used in the paper by Bahr (1992,1997) to model electrical anisotropy in real rocks. Bahr introduced the statistical dimensionless parameter “connectivity” to describe the influence of grain and liquid phase geometry on bulk conductivity of solid-liquid rock mixtures that can depend on the direction of EM wave propagation.

An analysis of the main results of the previous theoretical investigations of electrical and elastic properties of solid-liquid rock mixtures demonstrated that theoretical modelling of rock physical properties is a powerful tool to achieve a fundamental understanding of the main factors that affect both elastic and electrical parameters of rocks in the lithosphere. These main factors are the amount of liquid phase, rock microgeometry, pressure and temperature. They affect elastic and electrical properties in different manner, consequently, only simultaneous modelling of both elastic and electrical properties can give the answer to the question how they are connected.

It was not so easy, indeed, to find a universal model capable to describe both elastic and electrical properties of solid-liquid rock mixtures for the purpose of the present investigation, as theoretical rock models have been developed separately for elastic and electrical properties for a long time. The fractal rock model proposed in the Paper 4 is a combination of two theoretical rock models developed separately by Spangenberg (1998) for calculation of elastic properties of porous rocks and by Bahr (1997) for calculation of electrical conductivity of solid-liquid rock mixtures. The advantage of the fractal rock model is that it allows to model both isotropic and anisotropic elastic and electrical properties for rocks with different geometry of grain and liquid phase and gives estimates that agree well with those obtained by previously developed theoretical models.

3.3 Fractal model of solid-liquid rock system

The fractal model of porous rock was originally developed by Spangenberg (1998) for sedimentary rocks. The model belongs to a class of rock models with various

microstructure and allows to investigate the influence of rock microstructure on elastic properties. The main difference of the model from the other structural rock models is the use of a fractal approach to model the grain-to-grain contacts. The real rock is described by three main elements, i.e. matrix material, pore canals and contact region (Fig. 1 in Paper 4). The grain and pore size (l_i and a_i, respectively, i=*1,2,3*) can differ in x_1, x_2 and x_3 directions. The grain-to-grain contact region can be filled by N generations of self-similar substructures. The basic model is then subdivided into rectangular components of matrix material, contact region and pore fill combined either on serial or parallel connections. Then the Voight-Reuss bounds of elastic moduli of the basic model can be calculated for the resulting equivalent network under the assumption of homogeneous stress (Reuss, 1929) or of homogeneous strain (Voight, 1910). If the contact region of the model is filled with self-similar structures, then the calculation of elastic moduli is an iterative process starting with the calculation of the moduli for the smallest substructure model. The moduli of the smallest substructure then give the moduli for the contact region of the models of the following substructure etc. A similar iterative procedure can be used to estimate porosity, density, internal surface and surface of grain-to-grain contact.

Spangenberg (1998) pointed out two possible ways to subdivide the basic model into rectangular components that result in two different sets of equations for the calculation of elastic moduli. They are referred to as "horizontal subdivision" and "vertical subdivision", respectively. He also demonstrated that the main dependencies of elastic moduli on geometry, porosity, contact properties and rock composition are the same for both kinds of subdivision. That is why in the present work as well as in Paper 4 only relationships for the vertical subdivision of the basic model were used.

The plane wave modulus M and shear modulus μ of the smallest substructure of the model with vertical subdivision are:

$$M_i^v = \frac{a_j a_k M_p}{l_j l_k} + \frac{l_i M_p M_c \left((l_j - a_j)a_k + (l_k - a_k)a_j\right)}{l_j l_k \left(a_i M_c + (l_i - a_i)M_p\right)} + \frac{(l_j - a_j)(l_k - a_k) l_i M_c M_M}{l_j l_k \left((l_i - a_i)M_c + a_i M_M\right)}, \tag{3.11}$$

$$\mu_i^v = \frac{(l_j - a_j)(l_k - a_k) l_i \mu_c \mu_M}{l_j l_k \left((l_i - a_i)\mu_c + a_i \mu_M\right)}$$

where M_p, M_M, μ_p, μ_M are plane wave and shear moduli of pore and matrix materials, respectively, M_c and μ_c are the moduli of the contact area, respectively. P- and S-wave velocity are then calculated from eq. (3.4).

As Spangenberg (1998) pointed out, the model provides a high frequency velocity if the modulus of the pore fluid is substituted directly into the equations. The combination of the fractal rock model with Gassmann's (1951) theory makes it possible to estimate the low frequency velocity of fluid saturated rocks with different pore geometry. The elastic moduli of dry skeleton and porosity can be calculated for the basic model with fractal microstructure and then substituted into equations (3.2-3.4).

The contact conditions of the model are characterized by a contact parameter χ defined as a ratio of the solid contact faces and the surface of the base model.

$$\chi_i = \prod_{g=0}^{N} \frac{(l_{j,g} - a_{j,g})(l_{k,g} - a_{k,g})}{l_{j,g} l_{k,g}}, \text{where } i, j, k = 1,2,3 \qquad (3.12)$$

A similar fractal rock model was used by Bahr (1997) to investigate the effect of interconnected pores on electrical conductivity by Monte-Carlo simulation using the resistor random network approach and percolation theory. The bulk electrical conductivity σ_{eff} of a mixture of high conductivity material with conductivity σ_m and low conductivity material with conductivity σ_s can be presented as

$$\sigma_{eff} = \sigma_s + (2/3)\beta\sigma_m C(p), \qquad (3.13)$$

where *C(p)* is the connectivity that depends on the probability density *p* of certain pore geometry realisation, *C=1* corresponds to a perfectly interconnected high conducting phase.

Effective conductivity in the case when $\sigma_m >> \sigma_s$, i.e. for crustal rocks, is

$$\sigma_{eff} = (2/3)\beta\sigma_m C(p), \qquad (3.14)$$

It can be postulated that *C(p)* decreases with increasing contact region, i.e. the contact parameter in (3.12) defines the connectivity of the model: the larger it is, the smaller is the connectivity in (3.13-3.14). As the fractal contact parameter of the base model can vary in all three directions, the connectivity also can vary in all three directions, i.e.

$$C_i(p) = 1 - \chi_i \qquad (3.15)$$

Equations (3.11-3.15) can be used to calculate elastic constants and conductivity of rocks with fractal pore structure and describe rocks with isotropy as well as with transverse or orthorhombic anisotropy of physical properties. The results of Monte-Carlo simulations of elastic and electric properties of weakly porous crustal rocks and of partially molten rocks with the use of this model are presented in Paper 4.

The fractal rock model can also be used to demonstrate the combined effect of rock composition and pore microstructure on *Vp, Vs* and *Vp/Vs* ratio in the upper crust. For this purpose Monte-Carlo calculations of elastic properties of four selected rock samples with different mineral composition have been carried out. The rock samples from the KTB drill hole were described by Popp and Kern (1994). The main characteristics of the investigated rocks are given in Table 3.1. The elastic constants and intrinsic seismic velocity or rock matrix were calculated from modal mineralogy. No preferred mineral orientation was assumed. The modelling was performed for water saturated solid rock at

T=20^{0}C and P=600MPa. The elastic constants of pore content (saline water) were assumed to be M=2.3 GPa and μ=0. The grain and pore sizes were varied in all three directions, but they were restricted to ensure porosity less than 3%. Calculations were performed for 10000 different random pore configurations with different numbers of self-similar substructures in the contact area N. The simulation results are shown in Fig.3.1, Fig 3.2, Fig. 3.3 and Fig. 3.4. The apparent lines in the region of high contact parameter correspond to different number of sub-structures in the contact region N that is discrete. The case N=1 corresponds to fully isolated pores.

Table 3. Description of investigated rock samples (after Popp and Kern, 1994)

Sample	Rock description	Density (g/cm3)	Modal analysis (Vol %)	Intrinsic Vp (km/s)	Intrinsic Vs (km/s)	M (GPa)	m(GPa)
KTB 61C9b	Garnet Amphibolite	3.09	55%am, 17% plg., 10% gt., 18% acc.	6.52	3.83	131.36	45.33
KTB 99D5r	Biotite-Gneiss	2.75	42%qz, 27% plg, 24%bt, 4% ky, 3% gt	6.26	3.72	107.68	38.13
KTB H003B6a	Amphibolite/gneiss	2.84	45% fsp, 28% am, 15% qz, 12% acc.	6.42	3.81	117.05	41.23
KTB H014 G38a	Metagabbro	2.95	52%fsp, 36% am, 12% acc.	7.01	3.96	144.96	46.26

Abbreviations: fsp-feldspar, am-amphibole, gt-garnet, plg-plagioclase, qz-quartz, bt-biotite, ky-kyanite, acc-accessory minerals

The results demonstrate that the effect of microcracks on seismic macroproperties can be significant even in the case of low porosity. The ranges of possible values of seismic parameters of the samples KTB61C9b, KTBH003B6a and KTB99D5r overlap, although these rocks can be clearly discriminated by their density and intrinsic seismic velocity. It is interesting to note that microcracks have the strongest effect on *Vp/Vs* ratio.

Figure 2 demonstrates the results of Monte Carlo simulation of *Vp* for sample KTB99D5r in the high and low frequency ranges, i.e. with the use of fractal rock model and by combining of the fractal model with Gassmann's theory. The result demonstrates that the difference between the estimates given by the two methods is statistically insignificant.

a) b)

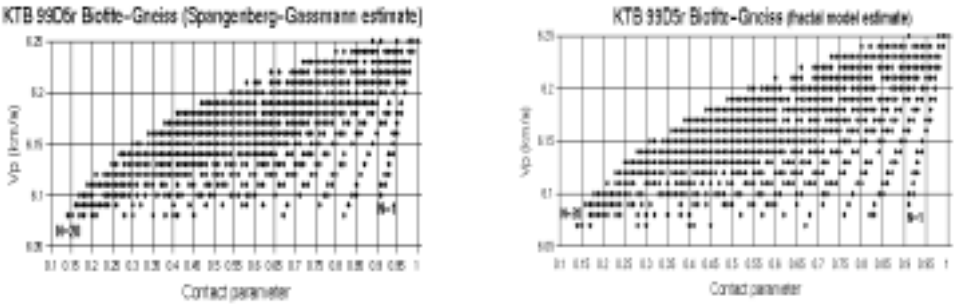

Fig. 2. Results of Monte-Carlo simulation of *Vp* of sample KTB99D5r for low porosity and various pore configurations; (a)- calculated from combination of fractal rock model and Gassmann's theory; (b)- calculated from fractal rock model. N denotes the number of substrucures in the contact area.

4 Inversion of magnetotelluric impedance tensor data by means of multiobjective optimisation

4.1 Problem formulation

The problem of magnetotelluric impedance tensor inversion with the aim of obtaining the 3-D resistivity structure of a region can be regarded as an example of a non-linear problem that cannot be effectively solved in the traditional Gaussian formulation. The practice of inverting the complex impedance tensor or apparent resistivity in the frequency domain instead of the EM fields in the time domain is based on the fact that the MT field observed on the surface is the sum of a primary field and a scattered secondary field caused by conductivity inhomogeneities. Using impedance instead of the spectra of the electric and magnetic components solves the problem of unknown amplitude of the primary field.

When the current encounters a region of discrete or permanent resistivity change, it sets up and maintains a charge distribution in the inhomogeneity region. These charges in turn produce their own electric fields, which must satisfy the boundary conditions (the current density normal to a boundary must be continuous across it). The electric field of the resulting charge distribution adds vectorially to the fields induced by the incident field. Thus the induced charges are responsible for the secondary EM fields of 2-D and 3-D structures.

When the incidence field is a plane wave, then the horizontal components of the EM field on the surfaces of non-homogeneous media are related by linear relationships:

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix} \begin{pmatrix} H_x \\ H_y \end{pmatrix}, \tag{4.1}$$

where $\mathbf{Z} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix}$ is the complex impedance tensor. In a uniform or horizontally layered Earth, $Z_{xx} = Z_{yy} = 0, Z_{yx} = -Z_{xy}$. In the case of a 2-D structure, if the x or y

axis is along the strike, then $Z_{xx} = Z_{yy} = 0$, but $Z_{xy} \neq Z_{yx}$. If neither axis is along strike, then $Z_{xx} = -Z_{yy} \neq 0$.

The case of 3-D structures is more complicated because of the wide range of possible 3-D objects, which can vary from small isolated conductivity anomalies to a regional anomalous distribution of conductivity. The effect of 3-D bodies on MT data has been studied by many authors, among others Wannamaker et al. (1984), who showed the effects of boundary charges on the MT apparent resistivity tensor, employing a MT theory for 3-D bodies in a layered Earth based on the assumption of a plane incident wave and 3-D inhomogeneity buried in the layered host. They obtained the following relationships for the components of the impedance tensor at low frequencies when the wavelengths in the host layers are long compared with the observation distance from the body:

$$\begin{aligned} Z_{xx} &\cong \left|Z_l\right|^2 \cdot \left|P_{xy}^0\right|^2, \\ Z_{xy} &\cong \left|Z_l\right|^2 \cdot \left|1 + P_{xx}^0\right|^2, \\ Z_{yx} &\cong \left|Z_l\right|^2 \cdot \left|1 + P_{yy}^0\right|^2, \\ Z_{yy} &\cong \left|Z_l\right|^2 \cdot \left|P_{yx}^0\right|^2, \end{aligned} \tag{4.2}$$

where $\left(\tilde{\mathbf{Z}}_\mathbf{l}\right) = \begin{pmatrix} 0 & Z_l \\ Z_l & 0 \end{pmatrix}$ is the layered Earth impedance and $P_{xx}^0, P_{xy}^0, P_{yx}^0, P_{yy}^0$ are components of a 3x2 normalised tensor $\left(\tilde{\mathbf{P}}_\mathbf{s}^\mathbf{0}(\mathbf{r})\right)$ representing the scattered field unique to a specified 3-D body, layered host and frequency:

$$\left(\tilde{\mathbf{P}}_\mathbf{s}^\mathbf{0}(\mathbf{r})\right) = \begin{pmatrix} \tilde{\mathbf{P}}_\mathbf{hs}^\mathbf{0}(\mathbf{r}) \\ \tilde{\mathbf{P}}_\mathbf{vs}^\mathbf{0}(\mathbf{r}) \end{pmatrix} = \begin{pmatrix} P_{xx}^0 & P_{xy}^0 \\ P_{yx}^0 & P_{yy}^0 \\ P_{zx}^0 & P_{zy}^0 \end{pmatrix}. \tag{4.3}$$

The scattered field tensors in (4.3) are subdivided into horizontal and vertical sub-tensors, denoted by the indices **hs** and **vs**, respectively.

Equations (4.2-4.3) demonstrate that all the components of the impedance tensor at low frequencies are distorted by boundary charge effects and are related to the impedance of a layered host by positive constants. The values of Z_{xy} and Z_{yx} can be equal to, less than or larger than Z_l, depending on the sign of P_{xx} or P_{yy}. Since $P_{ij}^{\ 0}$ become real as the frequency approaches zero (Wannamaker et al., 1984), the phases of all impedance components are asymptotic to that of the layered host medium. The diagonal elements of the impedance tensor depend mainly on the 3-D structure, but non-diagonal elements are also affected by the 2-D layered model. Consequently, it is necessary to use all the

components of the complex impedance tensor in the inversion procedure in order to reconstruct the 3-D resistivity distribution. Using only one component (Z_{xy}, for example) in the inversion procedure will increase the non-uniqueness of the inverse problem, because not all the information about the 3-D structure is taken into consideration.

4.2 Why the traditional formulation of the inverse problem cannot be applied to inversion of impedance tensor data

Let a 3-D resistivity structure within the Earth be described by a parameter vector $\mathbf{m}=[m_1,m_2,\ldots,m_k]\in A\subset M$, where M is the parameter space and A is a set of feasible solutions. Let the *a-priori* information on the unknown resistivity structure be expressed as a PDF $p(\mathbf{m})$. The new information on the vector $\mathbf{m}$ can be obtained by inverting data consisting of eight frequency-dependent components of the complex impedance tensor $\mathbf{Z}=\begin{bmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{bmatrix}$ at N observation points on the surface. These form the data vector $\mathbf{d}^{obs}=(d_1,d_2,\ldots,d_L)$. The complex impedance tensor is calculated from the observed time series of 4 components of the EM field. Maximisation of the *a-posteriori* conditional PDF of the vector $\mathbf{m}$, expressed in standard form as (1.10), gives an estimate for the parameter vector $\mathbf{m}$.

Under the assumption of Gaussian PDFs, maximisation of (1.10) is equivalent to minimisation of the following objective function:

$$L(\mathbf{m})=\left[g(\mathbf{m})-\mathbf{d}^{obs}\right]^{\mathbf{T}}\mathbf{C}_D^{-1}\left[g(\mathbf{m})-\mathbf{d}^{obs}\right]+(\mathbf{m}-\mathbf{m}^0)^{\mathbf{T}}\mathbf{C}_M^{-1}(\mathbf{m}-\mathbf{m}^0), \qquad (4.4)$$

where g is an operator that allows calculation of theoretical values for the impedance tensor from known values of the model parameters, $\mathbf{m}^0$ is the *a-priori* model, $\mathbf{C}_M$ and $\mathbf{C}_D$ are covariance matrices of the model and observed data, respectively, and $|\mathbf{C}_M|$ and $|\mathbf{C}_D|$ are the corresponding determinants. The first term in (4.4) is the difference between the observed and calculated model data in a weighted l_2 norm, and the second term introduces the *a-priori* knowledge about model parameters into the inversion procedure.

As was demonstrated in Chapter 1, there is no difference in principle between problem (4.4) and the Tikhonov regularisation scheme that was used recently in the procedure of 3-D MT impedance inversion proposed by Newman and Alumbaugh (2000). The Tikhonov regularisation scheme explicitly assumes Gaussian PDFs for the model parameters and experimental data, an assumption which is not valid when one has to invert the impedance tensor for a 3-D structure, because this cannot be regarded as the observed data vector. The components of the complex impedance tensor in the frequency domain are estimated from the experimental data observed in the time domain by solving the equation system (4.1), i.e. the impedance tensor is an estimate.

As pointed out by Menke (1989), using estimates instead of observations introduces a kind of distribution that differs from that of the original data. The distribution depends on the procedure used to calculate auto-spectra and cross-spectra from the observed values of the EM field.

It is known that standard least square estimation of the impedance tensor, which assumes a Gaussian PDF for the observed data, is highly unstable in the presence of outliers. These are a serious problem for spectral estimation, even with long data series (Jones et al., 1989). As Tarantola (1987) proposed, long-tailed PDFs should be used to model uncertainty in experimental data if outliers are suspected. The maximum likelihood function derived from the long-tail symmetric exponential PDF $p(\boldsymbol{x})=exp(-|\boldsymbol{x}|)$ corresponds to the minimum l_1 norm criterion that is known to be sufficiently insensitive to outliers (robust).

Jones et al. (1989) pointed out that there exist two types of error in magnetotelluric impedance estimation: statistical errors due to non-Gaussian effects (outliers and finite data) and bias errors due to noise power. They demonstrated that statistical errors can be removed, in principle, by applying robust estimates or by analyzing more data. The remote reference technology proposed by Goubau et al. (1978) and robust methods such as l_1 norm estimation and *M*-estimates (Chave and Thomson, 1989) can be used to avoid the problem of outliers in data. Biases cannot be fully removed even by robust methods, however, so that coherent noise sources, for example, can produce biases in remote reference processing. There exists an extensive literature on problems concerning estimation of the magnetotelluric impedance tensor, a full review of which would lie outside the scope of the present work.

One additional source of error exists in the problem of 3-D MT data inversion that is very seldom taken into consideration. A model EM field simulated with an integral equation or differential equation code may contain different types of numerical errors. The model impedance tensor components are then calculated from the model surface EM field. Mackie and Madden (1993) proposed the following relationships in order to calculate the model impedance tensor components from an EM field for two orthogonal source polarisations:

$$\begin{aligned} Z_{xx} &= \frac{E_{x1}H_{y2} - E_{x2}H_{y1}}{H_{x1}H_{y2} - H_{x2}H_{y1}}, \quad Z_{xy} = \frac{E_{x2}H_{x1} - E_{x1}H_{x2}}{H_{x1}H_{y2} - H_{x2}H_{y1}}, \\ Z_{xx} &= \frac{E_{y1}H_{y2} - E_{y2}H_{y1}}{H_{x1}H_{y2} - H_{x2}H_{y1}}, \quad Z_{xx} = \frac{E_{y2}H_{x1} - E_{y1}H_{x2}}{H_{x1}H_{y2} - H_{x2}H_{y1}}. \end{aligned} \tag{4.5}$$

If the components of the calculated EM field contain numerical simulation errors, the resulting distribution of the calculated impedance tensor components will not be Gaussian, as the errors will not be additive with respect to the impedance tensor components.

There also exist a number of additional sources of uncertainty in the problem of MT impedance tensor inversion that can affect the resolution of the inversion procedure, namely:

a) The 3-D resistivity structure has different effects on the components of the impedance tensor, diagonal elements being more sensitive to this structure than non-diagonal ones.

b) The EM response depends on the geometry of the experiment, i.e. the location of the stations with respect to the 3-D object. The responses at different observation points will have different sensitivities to the variation in model parameters.
c) The MT response is frequency dependent, i.e. the resolution decreases with depth.

It is obviously impossible to take the influence of all these factors into consideration in one common covariance matrix $\mathbf{C}_D$ in eq. (4.4).

It is seen from the above that the traditional Gaussian inversion scheme with a l_2 norm cannot be applied to invert the impedance tensor for a 3-D structure, even if robust methods are used to calculate the impedance from the experimental data. The other important conclusion is that standard testing of the inversion algorithm with synthetic data simulated by adding random noise to the impedance tensor components calculated for the model does not represent the real error distribution in the problem. This distribution can differ from a Gaussian one only due to non-additive errors produced by the forward problem calculation algorithm.

As has already been mentioned in Section 1.4, one more problem in the traditional inversion procedure is connected with proper formulation of the *a-priori* information. It is known that such information in inversion algorithms based on Tichonov regularisation is often formulated according to the requirements that the conductivity distribution must be the smoothest one, or the simplest one. Although the *a-priori* information in such a formulation helps to regularise the solution, the assumption of smoothness or simplicity is not valid in the case of sharp resistivity contrast between background conductivity and an embedded 3-D structure. Moreover, one very often has more knowledge of the conductivity structure in the region, e.g. drilling data, laboratory measurements of rock conductivity etc. This *a-priori* information can be formulated, following the maximum entropy principle, as an *a-priori* PDF of the parameter vector. Another possible approach proposed in Paper I is formulation of the *a-priori* information via fuzzy sets.

Summarising the above discussion, we can formulate the following requirements for the procedure of 3-D MT impedance inversion:

a) It has to be non-sensitive to outliers caused not only by errors in the experimental impedance tensor, but also by noise produced by the forward problem calculation.
b) It has to take into consideration the different sensitivities of the components of the impedance tensor to the variation in model parameters. This sensitivity can also depend on the geometry of the experiment and on the frequency.
c) It must be able to include different types of *a-priori* knowledge about the model parameters in the inversion scheme.

An algorithm for a 3-D impedance tensor inversion that satisfies the above requirements is described in the next section.

4.3 Algorithm for MT complex impedance tensor inversion based on the ideal point method of multiobjective optimization

Consider a set of experimental MT data that consist of eight components of the complex impedance tensor calculated for *n* frequencies at *m* observation points.

Assume that the data errors at each observation point are independent. In this case we can assume that the combined error due to noisy data, outliers and numerical errors in the forward problem calculation of each component of the complex impedance tensor at each frequency has a long-tail distribution that can be approximated by an exponential PDF. The corresponding maximum likelihood functions are then:

$$\begin{cases} F_1^j(\mathbf{m}) = \left\| Z_{xx}^{\mathrm{Re}}(\varpi_j) - \hat{Z}_{xx}^{\mathrm{Re}}(\varpi_j,\mathbf{m}) \right\|_{l_1} = \sum_{i=1}^{m} \frac{1}{\sigma_1^{ij}} \left| Z_{xx_i}^{\mathrm{Re}}(\varpi_j) - \hat{Z}_{xx_i}^{\mathrm{Re}}(\varpi_j,\mathbf{m}) \right|, \\ F_2^j(\mathbf{m}) = \left\| Z_{xy}^{\mathrm{Re}}(\varpi_j) - \hat{Z}_{xy}^{\mathrm{Re}}(\varpi_j,\mathbf{m}) \right\|_{l_1} = \sum_{i=1}^{m} \frac{1}{\sigma_2^{ij}} \left| Z_{xy_i}^{\mathrm{Re}}(\varpi_j) - \hat{Z}_{xy_i}^{\mathrm{Re}}(\varpi_j,\mathbf{m}) \right|, \\ F_3^j(\mathbf{m}) = \left\| Z_{yx}^{\mathrm{Re}}(\varpi_j) - \hat{Z}_{yx}^{\mathrm{Re}}(\varpi_j,\mathbf{m}) \right\|_{l_1} = \sum_{i=1}^{m} \frac{1}{\sigma_3^{ij}} \left| Z_{yx_i}^{\mathrm{Re}}(\varpi_j) - \hat{Z}_{yx_i}^{\mathrm{Re}}(\varpi_j,\mathbf{m}) \right|, \\ F_4^j(\mathbf{m}) = \left\| Z_{yy}^{\mathrm{Re}}(\varpi_j) - \hat{Z}_{yy}^{\mathrm{Re}}(\varpi_j,\mathbf{m}) \right\|_{l_1} = \sum_{i=1}^{m} \frac{1}{\sigma_4^{ij}} \left| Z_{yy_i}^{\mathrm{Re}}(\varpi_j) - \hat{Z}_{yy_i}^{\mathrm{Re}}(\varpi_j,\mathbf{m}) \right|, \\ F_5^j(\mathbf{m}) = \left\| Z_{xx}^{\mathrm{Im}}(\varpi_j) - \hat{Z}_{xx}^{\mathrm{Im}}(\varpi_j,\mathbf{m}) \right\|_{l_1} = \sum_{i=1}^{m} \frac{1}{\sigma_5^{ij}} \left| Z_{xx_i}^{\mathrm{Im}}(\varpi_j) - \hat{Z}_{xx_i}^{\mathrm{Im}}(\varpi_j,\mathbf{m}) \right|, \\ F_6^j(\mathbf{m}) = \left\| Z_{xy}^{\mathrm{Im}}(\varpi_j) - \hat{Z}_{xy}^{\mathrm{Im}}(\varpi_j,\mathbf{m}) \right\|_{l_1} = \sum_{i=1}^{m} \frac{1}{\sigma_6^{ij}} \left| Z_{xx_i}^{\mathrm{Im}}(\varpi_j) - \hat{Z}_{xx_i}^{\mathrm{Im}}(\varpi_j,\mathbf{m}) \right|, \\ F_7^j(\mathbf{m}) = \left\| Z_{yx}^{\mathrm{Im}}(\varpi_j) - \hat{Z}_{yx}^{\mathrm{Im}}(\varpi_j,\mathbf{m}) \right\|_{l_1} = \sum_{i=1}^{m} \frac{1}{\sigma_7^{ij}} \left| Z_{xx_i}^{\mathrm{Re}}(\varpi_j) - \hat{Z}_{xx_i}^{\mathrm{Re}}(\varpi_j,\mathbf{m}) \right|, \\ F_8^j(\mathbf{m}) = \left\| Z_{yy}^{\mathrm{Im}}(\varpi_j) - \hat{Z}_{yy}^{\mathrm{Im}}(\varpi_j,\mathbf{m}) \right\|_{l_1} = \sum_{i=1}^{m} \frac{1}{\sigma_8^{ij}} \left| Z_{xx_i}^{\mathrm{Re}}(\varpi_j) - \hat{Z}_{xx_i}^{\mathrm{Re}}(\varpi_j,\mathbf{m}) \right|, \quad j = 1,\ldots,n. \end{cases} \tag{4.6}$$

where Z and $\hat{Z}$ denote the impedances obtained from experimental data and calculated from the model parameter vector **m**, respectively (for simplicity, indices from 1 to 4 in the following correspond to real components of the complex impedance tensor and indices from 5 to 8 to its imaginary components).

We can then consider that the purpose of the inversion is to minimize the difference between the eight observed and calculated components of the complex impedance tensor in a l_1 norm for each frequency. In this case we have instead of one objective function a vector objective function with *n* x *8* components that maps the set of feasible solutions *A* from the parameter space into the vector objective space.

If we have a population of p points generated from the set of feasible solutions A, i.e. $B = \{\mathbf{m^1}, \mathbf{m^2}, ..., \mathbf{m^p}\}$, we can estimate maximum and minimum values for all the components of the vector misfit function on the set B, which form two vectors:

$$\mathbf{F}^{\max} = (F_1^{1\max}, F_2^{1\max}, ..., F_8^{n\max}), \quad \text{where } F_i^{j\max} = \max_{\mathbf{m}\in\mathbf{B}} F_i^j(\mathbf{m}), i = 1,...,8, \quad j = 1,...,n$$

and

$$\mathbf{F}^{\min} = (F_1^{1\min}, F_2^{1\min}, ..., F_8^{n\min}), \quad \text{where } F_i^{j\min} = \min_{\mathbf{m}\in\mathbf{B}} F_i^j(\mathbf{m}), i = 1,...,8, \quad j = 1,...,n.$$

It is then possible to define the ideal point in the objective space with the components:

$$\begin{cases} F_1^{j\,ideal} = \min\limits_{\mathbf{m}\in\mathbf{B}} F_1^j(\mathbf{m}), \\ F_2^{j\,ideal} = \min\limits_{\mathbf{m}\in\mathbf{B}} F_2^j(\mathbf{m}), \\ F_3^{j\,ideal} = \min\limits_{\mathbf{m}\in\mathbf{B}} F_3^j(\mathbf{m}), \\ F_4^{j\,ideal} = \min\limits_{\mathbf{m}\in\mathbf{B}} F_4^j(\mathbf{m}), \\ F_5^{j\,ideal} = \min\limits_{\mathbf{m}\in\mathbf{B}} F_5^j(\mathbf{m}), \\ F_6^{j\,ideal} = \min\limits_{\mathbf{m}\in\mathbf{B}} F_6^j(\mathbf{m}), \\ F_7^{j\,ideal} = \min\limits_{\mathbf{m}\in\mathbf{B}} F_7^j(\mathbf{m}), \\ F_8^{j\,ideal} = \min\limits_{\mathbf{m}\in\mathbf{B}} F_8^j(\mathbf{m}), \quad j = 1,...,n. \end{cases} \tag{4.7}$$

The estimate for the vector $\mathbf{m}$ in the parameter space will then be the point that minimises the distance between the vector objective function and the ideal point in the objective space in some norm:

$$\mathbf{m}^* = \min_{\mathbf{m}\in\mathbf{B}} \sum_{i=1}^{8} \sum_{j=1}^{n} \frac{|F_i^{j\,ideal} - F_i^j(\mathbf{m})|^k}{|F_i^{j\max} - F_i^{j\min}|^k}, \quad 1 \le k \le \infty. \tag{4.8}$$

The distance from the ideal point can be regarded as the maximum likelihood function of the generalised Gaussian PDF of order k of a random vector $\mathbf{F} = (F_1^1, F_2^1, ..., F_8^n)$ with the mean value $\mathbf{F}^{ideal}$ and with a variance approximately equal to $|F_i^{j\max} - F_i^{j\min}|^k/2$. If the noise in the impedance data is produced by shallow small-scale inhomogeneities, for example, then its effect will be stronger at high frequencies, i.e. the components of vector $\mathbf{F}$ corresponding to these frequencies can be regarded as outliers. In this case use of the l_1 norm in (4.8) can be recommended.

The normalising coefficient $|F_i^{j\max} - F_i^{j\min}|^k$ in (4.8) is an estimate of the range of variation in each component of the vector $\mathbf{F}$, i.e. the weights of the components of the vector objective function included in (4.8) are appointed in accordance with their sensitiveness to variation of the model parameters. It should be noted that the ideal point

estimated from the finite population of points is a pseudo-ideal, i.e. we need the number of points in the population p→∞ to find the "true" ideal point.
The following algorithm of MT impedance tensor inversion can then be formulated:

Step 1. Generate the initial population $B = \left\{\mathbf{m^1}, \mathbf{m^2}, ..., \mathbf{m^p}\right\}$ of models from the multidimensional parameter space in accordance with *a-priori* information on the model parameters. If the *a-priori* PDF of a vector **m** cannot be assumed to be Gaussian or uniform, the fuzzy set approach can be used to describe the *a-priori* information.

Step 2. Calculate values of the vector objective function (4.7) at each point in the initial population *B*.

Step 3. Estimate maximum and minimum values for each component of the vector objective function, i.e.

$$\mathbf{F}^{\max} = (F_1^{1\max}, F_2^{1\max}, ..., F_8^{n\max}), \quad \text{where } F_i^{j\max} = \max_{\mathbf{m}\in\mathbf{B}} F_i^j(\mathbf{m}), i = 1,...,8, \quad j = 1,...,n$$

and

$$\mathbf{F}^{\min} = (F_1^{1\min}, F_2^{1\min}, ..., F_8^{n\min}), \quad \text{where } F_i^{j\min} = \min_{\mathbf{m}\in\mathbf{B}} F_i^j(\mathbf{m}), i = 1,...,8, \quad j = 1,...,n.$$

Step 4. Find the ideal point in the objective space $\mathbf{F}^{ideal} = (F_1^{1ideal}, F_2^{1ideal}, ..., F_8^{nideal})$.

Step 5. Find an estimate for the vector **m** in the parameter space that minimises the distance between the vector objective function and the ideal point in the objective space (4.8).

Step 5 can be performed using the nearest neighbour algorithm of Sambridge (1998, 1999), a global optimisation algorithm that allows the information obtained in the previous steps to be use to reduce the number of forward problem calculations and concentrate the trial points in the area close to the global minimum. The initial population generated in Step 1 can be used for this purpose. The ideal point can be refined at this stage by taking account of information from the new trial points. A modification of the algorithm that allows trial points to be sampled in accordance with the membership function of a fuzzy set of feasible solutions is described in Paper I.

The proposed algorithm has been tested with synthetic magnetotelluric impedance tensor data, but its application and testing with real data remains a topic for future research.

5 Conclusions

The main targets of the joint interpretation of multimethod geophysical data are geological bodies with an established dependence between various physical properties (geophysical multiresponse). This relationship is probabilistic in nature and can be formally described by a function of some kind, e.g. a probability density function, regression equation etc. Thus, a detailed analysis of petrophysical data concerning rock properties that is aimed at defining this relationship is a necessary first step for the joint interpretation of multimethod geophysical data.

The parameters of geophysical multiresponse bodies can be estimated from joint inversion of the data, which can be effectively formulated within a Bayesian approach as a problem of maximisation of the joint *a-posteriori* probability density function of the observed data sets and *a-priori* information. Numerical optimisation of the joint *a-posteriori* PDF can be problematic due to the non-uniqueness that is always present in real geophysical inverse problems. Another problem is non-linearity, which can result from experimental data having a non-Gaussian PDF and a forward problem with a non-linear operator.

The effectiveness of joint inversion of multimethod geophysical data in a non-linear formulation can be analysed using the definitions of Kolmogorov's ε-entropy and the lower limit of resolution of the inversion procedure. This analysis demonstrates that the lower resolution limit of any procedure of joint inversion of multimethod geophysical data corresponds to the value of lower limit of resolution of the method with the best resolution, but cannot be less than this. The analysis also shows that joint inversion can reduce non-uniqueness in real geophysical inverse problems, so that the solution is better then that obtained from separate inversion of data from only one method.

The joint inversion of multimethod geophysical data can be regarded as a problem of multiobjective optimisation aimed at finding a set of Pareto optimal solutions. Multiobjective formulation allows the numerical methods of multiobjective optimisation to be applied for the purpose of geophysical data inversion and the development of computer algorithms capable of solving both linear and non-linear problems. An analysis of some well-known joint inversion techniques from the point of view of multiobjective optimisation theory demonstrates that they cannot be recommended in the case of non-linear inverse problems for which convexity is not guaranteed. Magnetotelluric

impedance tensor inversion for obtaining a 3-D resistivity structure is an example of a non-linear problem with observed data having a non-Gaussian PDF that cannot be effectively solved in a linear formulation. A robust algorithm for MT impedance tensor inversion can be constructed on the basis of the ideal point method of multiobjective optimisation.

The combination of various types of uncertain information into one inversion scheme can be an additional area of application for multiobjective optimisation techniques. Two types of uncertainty exist in geophysical inverse problems: the uncertainty attached to the observed data, which is probabilistic, as it results from random observation errors, and the uncertainty attached to the *a-priori* information, which can be called possibilistic, as it is connected with imperfect knowledge of the model parameters. This means that the use of measures of uncertainty other than probabilistic ones can be an efficient tool for the presentation of *a-priori* information in geophysical inverse problems. The two types of uncertainty can be combined into one common inversion scheme using the definition of Pareto optimality.

A necessary condition for the joint interpretation of seismic and gravity data is the well-known density-velocity relationship, which can also be used for parameterising the density model and transforming the non-linear gravimetric inverse problem into a linear one. Joint interpretation of gravity and seismic data by means of a density-velocity relationship demonstrated that bodies for which the density-velocity relationship properties are violated are often associated with important tectonic boundaries and can be targets of separate interpretation.

Statistical analysis of borehole data concerning velocity, density and electrical resistivity and the theoretical modelling of elastic and electrical properties of rocks allow the conclusion that solid-liquid rock mixtures can be a class of objects for which a quantitative relationship between electrical and elastic properties can be established. Theoretical modelling of the elastic and electrical properties of solid-liquid rock mixtures with a fractal microstructure has demonstrated the effect of rock microstructure on both elastic and electrical properties. This effect can reach significant values for water-saturated porous rocks in the upper crust and partially molten rocks in the mantle.

The main problem for the joint inversion of seismic and EM data is to separate the common factors affecting both seismic velocity and resistivity from factors that influence only elastic properties or only electrical properties. A multiscale presentation of the distribution of various physical properties can be used for this purpose. The geological medium under study can be presented as a superposition of random self-similar structures on different scales. The multiscale concept of a geological medium can provide the necessary theoretical background for the future development of methods for the joint inversion of multimethod geophysical data.

References

Anderson DL (1967) A seismic equation of state. Geoph Journal, 13: 9-30.

Archie GE (1942) The electrical resistivity log as an aid in determining some reservoir characteristics. Trans Americ Inst Mineral Met, 146: 54-62.

Babushka V and Cara M (1991) Seismic anisotropy in the Earth. Kluwer Academic Publishers, Dordrecht Boston London.

Bahr K (1992) Normale und fraktale Widerstandnetzwerke. In: Electromagnetische Tiefenforchung 14. Kolloquium. Borkheide, 25-29 May (Mai) 1992: 337-384.

Bahr K (1997) Electrical anisotropy and conductivity distribution functions of fractal random networks and of the crust: the scale effect of connectivity. Geoph J Int, 130: 649-660.

Barton PJ (1986) The relationship between seismic velocity and density in the continental crust- a useful constraint? Geoph J R Astron Soc, 87: 195-208.

Bean CJ (1996) On the cause of 1/f-power spectral scaling in borehole sonic logs. Geoph Res Letters, 23: 3119-3122.

Bellman RE and Zadeh LA (1970) Decision making in a fuzzy environment. Management science, 17: 141-164.

Berckhemler H, Rauen A, Winter H, Kern H, Kontny A, Lienert M, Nover G, Pohl J, Popp T, Schult A, Zinke J, Soffel HC (1997) Petrophysical properties of the 9-km-deep crustal section at KTB. J Geoph Res, 102, B8: 18,337-18,361.

Birch F (1961) The velocity of compressional waves in rocks to 10 kilobars (Part II). J Geoph Res, 65: 1083-1102.

Biot MA (1956a) Theory of propagation of elastic waves in a fluid satutated porous solid (I. Low frequency range). J Acoust Soc Am, 28, 2: 168-178.

Biot MA (1956b) Theory of propagation of elastic waves in a fluid satutated porous solid (II. Higher frequency range). J Acoust Soc Am, 28, 2: 179-191.

Bott MHP (1973) Inverse methods in the interpretation of magnetic and gravity anomalies. In: Bolt BA (ed) Methods in computational physics, Volume 13. Academic Press, New York: 123-162.

Bowman VJ (1976) On the relationship of the Tchebycheff norm and the efficient frontier of multi-criteria objectives. In: H Thiriez and S Zionts (eds) Multiple criteria decision making, Springer-Verlag, Berlin: 76-86.

Budiansky B and O'Connell RJ (1976) Elastic moduli of cracked solid. Int J Solids Structures, 12: 81-97.

Burke M and Fountain DM (1990) Seismic properties of rocks from an exposure of extended continental crust – new laboratory measurements from the Ivrea Zone. Tectonophysics, 182: 119-146.

Charnes A and Cooper WW (1960) Management models and industrial application of linear programming. Wiley, New York.

Chave AD and Thomson DJ (1989) Some comments on magnetotelluric response function estimation. J Geoph Res, 94, B10: 14,215-14,225.

Christensen NI and Mooney WD (1995) Seismic velocity and composition of the continental crust: a global view. J Geoph Res, 100, B7: 9761-9788.

Cooper RF (1986) Rhelogy and structure of olivine-basalt partial melt. J Geoph Res, 91, B9: 9315-9323.

DEKORP and Orogenic Processes Working Groups (1999) Structure of the Saxonian Granulites: geological and geophysical constraints on the exhumation of high-pressure/high-temperature rocks in the mid-European Variscan belt. Tectonics, 18, 5: 756-773.

Dubois D and Prade H (1997) Bayesian conditioning in possibility theory. Fuzzy sets and systems, 92: 223-240.

ELEKTB Group (1997) KTB and the electrical conductivity of the crust. J Geoph Res,102, B8: 18,289-18,305.

Elo S, Puranen R, Airo M (1978) Geological and areal variation of rock densities, and their relation to some gravity anomalies in Finland, GeoScrifter, 10: 123-164.

Emmermann R and Lauterjung J (1997) The German Continental Deep Drilling Program KTB: overview and major results. J Geoph Res,102, B8: 18,179-18,201.

Eshelby JD (1957) The determination of the elastic field of an ellipsoidal inclusion and related problems. Proc Roy Soc, Ser A, 221: 376-396.

EUROBRIDGE Seismic Working Group (1999) Seismic velocity structure across the Fennoscandia-Sarmatia suture of the East European Craton beneath the EUROBRIDGE profile through Lithuania and Belarus. Tectonophysics, 314: 193-217.

Faul U, Toomey D, Waff H (1994) Intergranular basaltic melt is distributed in thin, elongated inclusions. Geoph Res Lett, 21, 1: 29-32.

Gassmann F (1951) Über die Elastität poröser Medien. Vier Der Natur Gesellschaft, Zürich, 96: 1-22.

Gebrande H, Bopp M, Neurieder P, Schmidt T (1989) Crustal structure in the surrounding of the KTB Drill Site as devived from refraction and wide-angle seismic observations. In: Emmermann R and Wohlenberg J (eds) The German Continental Deep Drilling Program (KTB), Springer-Verlag, Berlin Heidelberg: 151-176.

Glaznev VN, Raevsky AB, Skopenko GB (1996) A three-dimensional integrated density and thermal model of the Fennoscandian lithosphere. Tectonophysics, 258: 15-33.

Goff JA and Holliger K (1999) Nature and origin of upper crustal seismic velocity fluctuations and associated scaling properties: combined stochastic analyses of KTB velocity and lithology logs. J Geoph Res, 104, B6: 13,169-13,182.

Goltsman MF (1982) Physical experiment and statistical conclusions. St Petersburg University. (in Russian)

Goubau WM, Gamble TD, Clarke J (1978) Magnetotelluric data analysis: Removal of bias. Geophysics, 43, 5: 1157-1162.

Grant FS and West GF (1965) Interpretation theory in applied geophysics. McGraw-Hill, New York.

Haimes YY, Ladson L, Wismer D (1971) On a bicriteria formulation of the problems of integrated system identification and system optimization. IEEE Transactions on Systems, Man and Cybernetics, SMC-1: 296-297.

Hashin Z and Shtrikman S (1963) A variational approach to the theory of the elastic behaviour of multiphase materials. J Mech Phys Solids, 11: 127-140.

Hartley RVL (1928) Transmission and information. The Bell System Technical Journal, 7: 535-563.

Hermance JF (1979) The electrical conductivity of materials containing partial melt: a simple model from Archie's law. Geoph Res Lett, 6, 7: 613-616.

Hjelt SE (1992) Pragmatic inversion of geophysical data. Springer-Verlag, Berlin Heidelberg.

Holliger K (1996) Upper-crustal seismic velocity heterogeneity as derived from a variety of P-wave sonic logs. Geoph J Int,125: 813-829.

Hurst H (1951) Long-term storage capacity of reservoirs, Trans Am Soc Civil Eng, 116: 770-808.

Jones AG, Chave AD, Egbert G, Auld D, Bahr K (1989) A comparison of techniques for magnetotelluric responce function estimation. J Geoph Res, 94: 14,201-14,213.

Jones AG and Holliger K (1997) Spectral analysis of the KTB sonic and density logs using robust nonparametric methods. J Geoph Res, 102, B8: 18,391-18,403.

Jones T D and Nur A (1984) The nature of seismic reflections from deep crustal fault zones. J Geoph Res, 89, B5: 3153-3171.

Karatayev GI and Pashkevich IK (1986) Geological and mathematical analysis of sets of several geophysical methods data. Navukova Dumka, Kiev. (in Russian)

Karatayev GI and Kozlovskaya E (1997) Method of integrated computer interpretation of DSS, gravity and magnetic data and its application to density modeling of the Earth's crust and upper mantle. Litosphere, 6: 117-128.

Kern H and Wenk HR (1990) Fabric-related velocity anisotropy and shear wave splitting in rocks from Santa Rosa mylonite zone, Calofornia. J Geoph Res, 95: 11 213-11 223.

Kern H, Schmidt R, Popp T (1991) The velocity and density structure of the 400 m crustal segment at the KTB drilling site and their relationship to lithological and microstructural characteristics of the rock: an experimental approach. Scientific Drilling, 2: 130-145.

Khalevin NN, Aleinikov AL, Kolupajeva EN, Tiunova AM, Yunusov FF (1986) On the joint use of longitudinal and transverse waves in deep seismic sounding. Geologia i Geofizika, 10: 94-98. (in Russian)

Klir GJ (1990) A principle of uncertainty and information invariance. International Journal of General Systems, 17, 2-3: 249-275.

Klir GJ and Yuan B (1995) Fuzzy sets and fuzzy logic. Theory and applications. Prentice-Hall RTR, New Jersey.

Kneib G (1995) The statistical nature of the upper continental crystalline crust derived from in situ seismic measurements. Geoph J Int, 122: 594-616.

Kolmogorov AN and Tichomirov VM (1959) ε-enthropy and ε-cardinality of sets in functional spaces. Uspechi matematicheskich nauk, 14, 2(86): 4-61.(in Russian)

Kozlovskaya E, Karatayev GI, Yliniemi J, Giese R (1999) Deep structure of the Earth's crust and upper mantle along EUROBRIDGE-96 DSS profile: results from integrated interpretation of seismic and gravity data. In: 7th EUROBRIDGE Workshop: Between EUROBRIDGE and TTZ. Suvalki-Szelment, Poland: 51-55.

Kozlovskaya E, Yliniemi J (1999) Integrated seismic and density model of the Earth's crust and upper mantle beneath the LT-7 and TTZ DSS profiles, Poland. In: 61st EAGE Conference and Technical Exhibition. Helsinki, Finland, Extended abstract book: P145.

Krasovsky SS (1981) Reflection of continental-type crustal dynamics in the gravity field. Navukova Dumka. Kiev. (in Russian)

Kuhn HW and Tucker AW (1951) Nonlinear programming. In: Neyman J (ed) Proceedings of the Second Berkley Symposium on Mathematical Statistics and Probability. University of California Press: 481-492.

Lamoureux G, Ildefonse B, Mainprice D (1999) Modelling of seismic properties of fast-spreading ridge crustal low-velocity zones: insights from Oman gabbro textures. Tectonophysics, 312: 283-301.

Leary P (1991) Deep borehole rock evidence for fractal distribution of fractures in crystalline rocks. Geoph J Int, 107: 615-627.

Ludvig JF, Nafe JE, Drake CL (1970) Seismic refraction. In: Maxwell A E (ed) The sea. Vol 4, Willey, New York: 53-84.

Mackie RL, Madden TR, Wannamaker PE (1993) Three-dimensional magnetotelluric modelling using difference equations- theory and comparison to integral equations solutions, Geophysics, 58: 215-226.

Mainprice D (1997) Modelling the anisotropic seismic properties of partially molten rocks found at Mid-Ocean Ridges. Tectonophysics, 279: 161-179.

Madden TR (1976) Random networks and mixing laws. Geophysics, 41: 1104-1125.

Mandelbrot B (1983) The Fractal Geometry of Nature. Freeman, San-Francisco.

Mandelbrot B (1989) Multifractal measures, especially for the geophysicist. Pure Appl Geoph, 131, 1/2: 5-42.

Marsan D and Bean CJ (1999) Multiscaling nature of sonic velocities and lithology in the upper crystalline crust: evidence from the KTB Main Borehole. Geoph Res Lett, 26, 2: 275-278.
Mavko GM and Nur A (1978) The effect of nonelliptical cracks on the compressibility of rocks, J Geoph Res, 83, B9: 4459-4468.
Mavko GM (1980) Velocity and attenuation in partially molten rocks. J Geoph Res, 85, B10: 5173-5189.
Meissner R and Rabbel W (1999) Nature of crustal reflectivity along the DEKORP profiles in Germany in comparison with reflection patterns from different tectonic units worldwide: A review. Pure Appl Geoph, 156: 7-28.
Menke W (1989) Geophysical data analysis: Discrete inverse theory. Academic Press, London.
Morozova EA and Pavlenkova NI (1995) Ambiguity problem in construction a seismic crustal model for the southeastern Barents Sea. Physics of the Solid Earth (English translation), 31,2: 164-174.
Nafe HE and Drake CL (1963) Physical properties of marine sediments. In: The Sea. Vol 3. Interscience: 794-815.
Newman G A and Alumbaugh D L (2000) Three-dimensional magnetotelluric inversion using non-linear conjugate gradients. Geoph J Int, 140: 410-424.
Novikov PS (1937) On the uniqueness of the solution of inverse problem of the potential. DAN USSR, 115: 1. (in Russian)
Peching R, Haverkamp R, Wohlenberg J, Zimmermann G, Burkhardt H (1997) Integrated log interpretation in the German Continental Deep Drilling program: Lithology, porosity and fracture zones. J Geoph Res, 102, B8: 18,363-18,390.
Parker RL (1977) Understanding inverse theory. Ann Rev Earth Planet Sci,5: 35-64.
Podinovsky VV and Nogin VD (1982) Pareto-optimal solutions of multicriteria problems. Nauka, Moscow. (in Russian)
Popp T and Kern H (1994) The influence of dry and water saturated cracks on seismic velocities of crustal rocks – a comparison of experimental data with theoretical model. Surveys in Geophysics, 15, 5: 443-466.
Pugachev VS (1965) Theory of random functions and its application to control problems. Pergamon Press, Oxford.
Rabbel W (1994) Seismic anisotropy at the Continental Deep Drilling Site (Germany). Tectonophysics, 232: 329-341.
Rabbel W and Mooney WD (1996) Seismic anisotropy of the crystalline crust: what does it tell us? Terra Nova, 8: 16-21.
Rabbel W, Siegesmund S, Weiss T, Pohl M, Bohlen T (1998) Shear wave anisotropy of laminated lower crust bemneath Urach (SW Germany): a comparison with xenoliths and with exposed lower crustal sections. Tectonophysics, 298: 337-356.
Reuss A (1929) Berechnung der Fliergenze von Mischkristallen. Angew Mathem und Mech, 9: 49-58.
Rink M and Schopper JR (1968) Computation of network models of porous media. Geoph Prosp, 16: 277-294.
Roberts JJ and Tyburczy JA (1991) Frequency dependent electrical properties of polycrystalline olivine compacts. J Geoph Res, 96, B10: 16,205-16,222.
Sakawa M and Yano H (1990) Trade-off rates in the hyperplane method for multiobjective optimization problems. European Journal of Operational Research, 44: 105-118.
Sakawa M (1993) Fuzzy sets and interactive multiobjective optimization. Plentum Press, New York.
Sambridge M (1998) Exploring multidimensional landscapes without a map. Inverse problems, 14: 427-440.
Sambridge M (1999) Geophysical inversion with a neighbourhood algorithm-I. Searching a parameter space. Geoph J Int, 138: 479-494.
Schmeling H (1985) Numerical models on the influence of partial melt on elastic, anelastic and electrical properties of rocks. Part I: elasticity and anelasticity. Phys Earth Planet Int, 41: 34-57.
Schmeling H (1986) Numerical models on the influence of partial melt on elastic, anelastic and electrical properties of rocks. Part II: electrical conductivity. Phys Earth Planet Int, 43: 123-136.

Schön JH (1983) Petrophysik. Akademie Verlag Berlin and Ferd. Enke Verlag Stuttgart.
Schön JH (1998) Physical properties of rocks: fundamentals and principles of petrophysics. Pergamon.
Shafer G (1976) A mathematical theory of evidence. Princeton University Press.
Shannon CE (1948) The mathematical theory of communication. The Bell System Technical Journal, 27, 379-423: 623-656.
Skeels DC (1947) Ambiguity in gravity interpretation. Geophysics, 12: 43-56.
Spangenberg E (1998) A fractal model for physical properties of porous rock: Theoretical formulations and application to elastic properties. J Geoph Res, 103, B6: 12,269-12,289.
Sobol IM (1985). Multicriteria interpretation of method of regularization of ill-posed problems. Preprint of Institute of Applied Mathematics. Academy of Sciences of the USSR, Moscow. (in Russian)
Sobolev SV and Babeyko AY (1994) Modeling of mineralogical composition, density and elastic wave velocities in anhydrous magmatic rocks. Surveys in Geophysics, 15: 515-544.
Sugeno M (1977) Fuzzy measures and fuzzy integrals- A survey. In: Gupta MM, Saridis GN, Gaines BR (eds) Fuzzy automata and decision process. Elsevier, Amsterdam, New York.
Tarantola A (1987) Inverse problem theory. Elsevier, Amsterdam, New York.
Tikhonov AN and Arsenin VY (1977) Solution of ill-posed problems. VH Winston &Sons, Washington, DC.
Toksös MN, Cheng CH, Timur A (1976) Velocities of seismic waves in porous rocks, Geophysics, 41, 4: 621-642.
Voigt W (1910) Lerhbuch der Kristallphysic, Teubner-Verlag, Leipzig.
Waff HS (1974) Theoretical considerations of electrical conductivity in a partially molten mantle and implications for geothermometry. J Geoph Res, 79: 4003-4010.
Walsh JB (1965) The effect of cracks on the Poisson ratio, J Geoph Res, 70, 20: 5249-5257.
Wang Z and Klir GJ (1992) Fuzzy measure theory. Plentum Press, New York.
Wannamaker PE, Hohmann GW, Ward SH (1984) Magnetotelluric responses of three-dimentional bodies in layered Earth. Geophysics, 49, 9: 1517-1533.
Watanabe T and Kurita K (1993) The relationship between electrical conductivity and melt fraction in a partially molten simple system: Archie's law behaviour. Phys Earth Plan Int, 78: 9-17.
White J E (1983) Underground sound, Application of seismic waves. Methods Geochem Geophys Ser, Vol 18, Elsevier, NewYork.
Wu RS, Xu Z, Li XP (1994) Heterogeneity spectrum and scale-anisotropy in the upper crust revealed by the German Continental Deep Drilling (KTB) Holes. Geoph Res Lett, 21: 911-914.
Yager RR (1984) A representation of the probability of fuzzy subsets. Fuzzy sets and systems, 13: 273-278.
Yu PL (1973) A class of solutions for group decision problems, Management Science, 19: 936-946.
Zadeh LA (1965) Fuzzy sets. Information and control, 12, 2: 94-102.
Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy sets and systems, 1: 3-28.
Zeleny M (1973) Compromise programming. In: Chochrane JL and Zeleny M (eds) Multiple criteria decision making. University of South California Press: 262-301.
Zeleny M (1982) Multiple criteria desision making. McGraw-Hill, New York.
Zidarov D (1990) Inverse gravimetric problem in geoprospecting and geodesy. Elsevier, Amsterdam.
Zimmermann HJ (1978) Fuzzy programming and linear programming with several objective functions. Fuzzy sets and systems, 1: 45-55.

Appendix 1

Fuzzy measures

The concept of a fuzzy measure was introduced by Sugeno (1977) and later developed by Wang and Klir (1992). It provides a broad framework for understanding the difference between various measures of uncertainty.

Definition A1.1 (Klir and Yuan, 1995). Given a universal set X and a non-empty family P of subsets of *X, a fuzzy measure* on $\langle X, \mathsf{P} \rangle$ is a function $g : \mathsf{P} \to [0,1]$ that satisfies the following requirements:

d) $g(\varnothing) = 0$ and $g(X) = 1$ (boundary requirements);

e) for all $A, B \in \mathsf{P}$, if $A \subseteq B$, then $g(A) \le g(B)$ (monotonicity);

f) for any increasing sequence $A_1 \subset A_2 \subset ... \text{ in } \mathsf{P}$, if $\bigcup_{i=1}^{\infty} A_i \in \mathsf{P}$, then $\lim_{i \to \infty} g(A_i) = g(\bigcup_{i=1}^{\infty} A_i)$ (continuity from below);

g) for any decreasing sequence $A_1 \supset A_2 \supset ... \text{ in } \mathsf{P}$, if $\bigcup_{i=1}^{\infty} A_i \in \mathsf{P}$, then $\lim_{i \to \infty} g(A_i) = g(\bigcup_{i=1}^{\infty} A_i)$ (continuity from above).

Requirements (c-d) are applicable only to an infinite universal set. Fuzzy measures as defined by *Definition A1.1* are a generalisation of probability measures obtained by replacing the additivity requirement by the weaker requirements of monotonicity and continuity.

The number *g(A)* assigned to a set *A* by a fuzzy measure *g* signifies the total available evidence that a given element of *X* whose characterisation is deficient in some respect belongs to *A*.

The evidence theory is based on dual non-additive measures, *belief measures* and *plausibility measures.*

Definition A1.2 Given a universal finite set X, a *belief measure* is a function

$Bel : \mathbb{Q}(X) \to [0,1]$ such that $Bel(\varnothing) = 0, Bel(X) = 1$ and

$$Bel(A_1 \cup A_2 \cup ... \cup A_n) \geq \sum_j Bel(A_j) - \sum_{j<k} Bel(A_j \cap A_k) + ... + (-1)^{n+1} Bel(A_1 \cap A_2 \cap ... \cap A_n \tag{A1.1}$$

for all possible families of subsets of X.
For each $A \in \mathbb{Q}(X)$, *Bel(A)* is interpreted as the degree of belief (based on available evidence) that a given element of *X* belongs to the set *A*. The subsets of *X* can be viewed as answers to particular questions.
When the sets $A_1, A_2, ..., A_n$ in (A1.1) are pairwise disjoint, the inequality requires that the degree of belief associated with the union of sets is not smaller than the sum of the degrees of belief pertaining to the individual sets. This basic property of belief measures is thus a weaker version of the additivity property of probability measures. This implies that probability measures are special cases of belief measures for which the equality in (A1.1) is always satisfied. The fundamental property of belief measures, which follows from (A1.1), is that

$$Bel(A) + Bel(\overline{A}) \leq 1. \tag{A1.2}$$

The *plausibility measure* associated with each belief measure is defined by the following equation:

$$Pl(A) = 1 - Bel(\overline{A}) \tag{A1.3}$$

for all $A \in \mathbb{Q}(X)$. Similarly,

$$Bel(A) = 1 - Pl(\overline{A}). \tag{A1.4}$$

Belief measures and plausibility measures are therefore mutually dual. Plausibility measures can also be defined independently of belief measures.
Definition 1.3. A *plausibility measure* is a function
$Pl : \mathbb{Q}(X) \to [0,1]$ such that $Pl(\varnothing) = 0, Pl(X) = 1$ and

$$Pl(A_1 \cap A_2 \cap ... \cap A_n) \leq \sum_j Pl(A_j) - \sum_{j<k} Pl(A_j \cup A_k) + ... + (-1)^{n+1} Pl(A_1 \cup A_2 \cup ... \cup A_n) \tag{A1.5}$$

for all possible families of subsets of *X*.
From (A1.5) it follows that

$$Pl(A) + Pl(\overline{A}) \geq 1. \tag{A1.6}$$

Belief and plausibility measures can conveniently be characterised by a function

$$m : \mathcal{Q}(X) \to [0,1] \text{ such that } m(\varnothing) = 0 \text{ and } \sum_{A \in \mathcal{Q}(X)} m(A) = 1 . \qquad \text{(A1.7)}$$

This function is called *a basic probability assignment*. For each set $A \in \mathcal{Q}(X)$, the value *m(A)* expresses the proportion to which all available and relevant evidence supports the statement that a particular element of X belongs to the set A. The value *m(A)* is a characteristic only of set A. If there is some additional evidence that the element belongs to a subset of A, i.e. $B \subseteq A$, it must be expressed by another value *m(B)*. The fundamental difference between probability distribution functions and basic probability assignments is that the former are defined on *X* while the latter are defined on $\mathcal{Q}(X)$ It is not required for basic probability assignments that *m(X)=1* or $m(A) \le m(B)$ when $A \subseteq B$. From this it follows that basic probability assignments are not fuzzy measures.
Given a basic probability assignment *m*, a *belief measure* and a *plausibility measure* are uniquely determined for all sets $A \in \mathcal{Q}(X)$ by the equations

$$\begin{aligned} Bel(A) &= \sum_{B \mid B \subseteq A} m(B), \\ Pl(A) &= \sum_{B \mid A \cap B \neq \varnothing} m(B). \end{aligned} \qquad \text{(A1.8)}$$

Equation (A1.8) means that while *m(A)* characterises the degree of evidence for the belief that some element examined belongs to the set *A* alone, *Bel(A)* represents the total evidence for the belief that the element in question belongs to set *A* as well as to various special subsets of *A*. The plausibility measure represents not only the total evidence that the element belongs to set *A* or to any other of its subsets, but also the additional belief associated with sets that overlap with *A*. From this it follows that

$$Pl(A) \ge Bel(A) \forall A \subset \mathcal{Q}(X) . \qquad \text{(A1.9)}$$

A basic probability assignment can be obtained from the corresponding belief and plausibility measures, i.e. each of the three functions *m, Bel* and *Pl* is sufficient to determine the other two.
Every set $A \in \mathcal{Q}(X)$ for which $m(A) > 0$ is usually called *a focal element of m*. The pair $\langle \mathcal{F}, m \rangle$ where $\mathcal{F}$ and *m* are a set of focal elements and associated basic assignments, respectively, is called *a body of evidence*.
Total ignorance is expressed in terms of a basic assignment by the statement $m(X) = 1$ and $m(A) = 0$ for all $A \neq X$. The expression of total ignorance in terms of a corresponding belief measure is the same: $Bel(X) = 1$ and $Bel(A) = 0$ for all $A \neq X$, but in terms of the corresponding plausibility measure it is different: $Pl(\varnothing) = 1$ and $Pl(A) = 1$ for all $A \neq \varnothing$.
A special branch of evidence theory that deals only with bodies of evidence whose focal elements are nested is referred to as *possibility theory*. Special counterparts of belief measures and plausibility measures in possibility theory are called *necessity measures*

and *possibility measures,* denoted as *Nec* and *Pos*, respectively. The general formulation of necessity and possibility measures is given by the following two definitions:

Definition A1.4: Let *Nec* denote a fuzzy measure on $\langle X, \mathbb{Q} \rangle$. Then, *Nec* is called a *necessity measure* if and only if $Nec\left(\bigcap_{k \in K} A_k\right) = \inf_{k \in K} Nec(A_k)$ for any family $\{A_k \mid k \in K\}$ in $\mathbb{Q}$ such that $\bigcap_{k \in K} A_k \in \mathbb{Q}$, where *K* is an arbitrary index set.

Definition A1.5: Let *Pos* denote a fuzzy measure on $\langle X, \mathbb{Q} \rangle$. Then, *Pos* is called a *possibility measure* if and only if $Pos\left(\bigcap_{k \in K} A_k\right) = \sup_{k \in K} Pos(A_k)$ for any family $\{A_k \mid k \in K\}$ in $\mathbb{Q}$ such that $\bigcap_{k \in K} A_k \in \mathbb{Q}$, where K is an arbitrary index set.

Since necessity and possibility measures are special cases of belief and plausibility measures, respectively, they satisfy conditions (A1.2-A1.4):

$$\begin{aligned} &Nec(A) + Nec(\overline{A}) \le 1, \\ &Pos(A) + Pos(\overline{A}) \ge 1, \\ &Nec(A) = 1 - Pos(\overline{A}). \end{aligned} \tag{A1.10}$$

It also follows from the definitions of *Pos* and *Nec* that

$$\begin{aligned} &\min\left[Nec(A), Nec(\overline{A})\right] = 0, \\ &\max\left[Pos(A), Pos(\overline{A})\right] = 1. \end{aligned} \tag{A1.11}$$

It can also be proved (Klir and Yuan, 1995) that for every $A \in \mathbb{Q}(X)$, any necessity measure, *Nec*, on $\mathbb{Q}(X)$ and the associated possibility measure, *Pos*, will satisfy the following conditions:

a) $Nec(A) > 0 \Rightarrow Pos(A) = 1$;

b) $Pos(A) > 1 \Rightarrow Nec(A) = 0$.

An important property of possibility theory is that every possibility measure on a finite power set $\mathbb{Q}(X)$ is uniquely represented by the associated *possibility distribution function* $r : X \to [0,1]$ via the formula

$$Pos(A) = \max_{x \in A} r(x) \tag{A1.12}$$

for each $A \in \mathbb{Q}(X)$. When *X* is not finite, (A1.12) must be replaced by the more general equation:

$$Pos(A) = \sup_{x \in A} r(x) . \qquad \text{(A1.13)}$$

Possibility theory can be formulated not only in terms of nested bodies of evidence but also in terms of *fuzzy sets* (see Appendix 2 for a fuzzy set definition), since fuzzy sets, like possibilistic bodies of evidence, are also based on families of nested sets, the appropriate α-cuts. Possibility measures are directly connected to fuzzy sets via the associated possibility distribution functions.

As postulated by Zadeh (1978), a *possibility* that $x = u$, $x \in X$ is equal to the membership function of a fuzzy set defined in *X*, i.e. the *possibility distribution function* $r_X(x)$ associated with *X* is equal to a membership function of a fuzzy subset *A* defined on *X,* namely $\mu_A(x)$ (see Appendix 2 for more detailed description of fuzzy set theory).

A *probability measure,* as a special class of fuzzy measures, must satisfy a requirement that is called *the additivity axiom of probability measures*: if $A_1, A_2, ...A_n$ is any sequence of pairwise disjoint sets in $\mathcal{Q}(X)$, then $Pro(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} Pro(A_i)$. This axiom is stronger than the superadditivity axiom of belief measures (A1.1), i.e. probability measures can be regarded as a special type of belief measures. The relationship between belief measures and probability measures can be characterised by the following theorem (proof to be found in Klir and Yuan, 1995):

Theorem A1.1. A belief measure *Bel* on a finite power set $\mathcal{Q}(X)$ is a probability measure if and only if the associated basic probability assignment function m is given by $m(\{x\}) = Bel(\{x\})$ and $m(A) = 0$ for all subsets in X that are not singletons.

The main difference between probabilities and possibilities lies in the structure of the respective bodies of evidence. While probabilistic bodies of evidence consist of singletons, possibilistic bodies of evidence are families of nested sets. While possibility theory is based on dual measures, i.e. special cases of belief and plausibility measures, probability theory deals with the case when belief measures and plausibility measures are equal.

Belief measures and plausibility measures may be interpreted as lower and upper probability estimates. In this interpretation, the dual measures Bel and Pl are used to form intervals which are viewed as imprecise estimates of probabilities.

Appendix 2

Basic definitions of a fuzzy set theory

Definition A2.1 (Zadeh, 1965): Let *X* denote a universal set. Then a *fuzzy subset A* of *X* is defined by its membership function $\mu_A : X \to [0,1]$, which assigns to each element $x \in X$ a real number $\mu_A(x)$ in the interval [0,1], where the value of $\mu_A(x)$ represents a grade of membership of *A* for *x*. Thus, the nearer the value of $\mu_A(x)$ is to unity, the higher the grade of membership of *A* that applies to *x*.

A *fuzzy subset A* can be characterized as a set of ordered pairs of an element x and a membership grade $\mu_A(x)$, and is often written $A = \{(x, \mu_A(x)) \mid x \in X\}$.

The following basic notions are defined for fuzzy sets (Sakawa, 1993).

Definition A2.2. The support of a fuzzy set A on X, denoted by supp(A), is the set of points in X at which $\mu_A(x)$ is positive, i.e.,

$$\text{supp}(A) = \{x \in X \mid \mu_A(x) > 0\}.$$

Definition A2.3. The height of a fuzzy set A on X, denoted by hgt(A), is the least upper bound of $\mu_A(x)$, i.e.,

$$\text{hgt}(A) = \sup_{x \in X} \mu_A(x).$$

Definition A2.4. A fuzzy set *A* on *X* is said to be *normal* if its height is unity, i.e., if there is $x \in X$ such that $\mu_A(x) = 1$. If this is not the case, the fuzzy set is said to be subnormal.

Definition A2.5. A fuzzy set A on x is *empty*, denoted by $\varnothing$, if and only if $\mu_A(x) = 0$ for all $x \in X$. The universal set *X* can obviously be viewed as a fuzzy set whose membership function is $\mu_X(x) = 1$.

Certain set theory operations with fuzzy sets originally proposed by Zadeh (1965) are defined as follows.

Definition A2.6. Equality: The fuzzy sets *A* and *B* on *X* are equal, denoted by $A = B$, if and only if their membership functions are equal everywhere on X:

$A = B \Leftrightarrow \mu_A(x) = \mu_B(x)$ for all $x \in X$.

Definition A2.7. Containment: The fuzzy set A is contained in B (or is a subset of B), denoted by $A \subseteq B$, if and only if its membership function is less than or equal to that of B everywhere on X:

$A \subseteq B \Leftrightarrow \mu_A(x) \leq \mu_B(x)$ for all $x \in X$.

Definition A2.8. Complementation: The complement of a fuzzy set A on X, denoted by $\overline{A}$, is defined by

$\mu_{\overline{A}}(x) = 1 - \mu_A(x)$ for all $x \in X$.

Definition A2.9. Intersection: The intersection of two fuzzy sets A and B on X, denoted by $A \cap B$, is defined by

$\mu_{A \cap B}(x) = \min\{\mu_A(x), \mu_B(x)\}$ for all $x \in X$.

Definition A2.10. Union: The union of two fuzzy sets A and B on X, denoted by $A \cup B$, is defined by

$\mu_{A \cup B}(x) = \max\{\mu_A(x), \mu_B(x)\}$ for all $x \in X$.

As Zadeh (1965) pointed out, it is possible to extend many of the basic properties which hold for ordinary sets to fuzzy sets. The following properties for union, intersection and complementation hold for fuzzy sets in a similar manner to ordinary sets.

Commutativity laws:

$A \cup B = B \cup A$, $A \cap B = B \cap A$.

Associativity laws:

$A \cup (B \cup C) = (A \cup B) \cup C$, $A \cap (B \cap C) = (A \cap B) \cap C$.

Distributivity laws:

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

De Morgan's laws:

$\overline{(A \cup B)} = \overline{A} \cap \overline{B}$, $\overline{(A \cup B)} = \overline{A} \cup \overline{B}$.

Involution:

$\overline{\overline{A}} = A$.

The concept of α-level sets or α-cuts is an important transfer between ordinary sets and fuzzy sets. It also plays an important role in the construction of a fuzzy set by a series of ordinary sets.

Definition A2.11. (Sakawa, 1993): The *α-level set* or *α-cut of a fuzzy set A* is defined as an ordinary set A_α for which the degree of the membership function exceeds the level α:

$A_\alpha = \{x \mid \mu_A(x) \geq \alpha\}, \alpha \in [0,1]$.

The following evident property holds for α-level sets:

$\alpha_1 \leq \alpha_2 \Leftrightarrow A_{\alpha_1} \supseteq A_{\alpha_2}$.

The following *decomposition theorem* can be proved for a fuzzy set (see Sakawa, 1993 for proof).

Theorem A2.1: A fuzzy set A can be represented by

$$A = \bigcup_{\alpha \in [0,1]} \alpha A_\alpha ,$$

where αA_α denotes the algebraic product of a scalar α with the α-level set A_α, i.e. its membership function (characteristic function) is given by

$\mu_{\alpha A_\alpha}(x) = \alpha \mu_{A_a}(x) = \alpha C_{A_\alpha}(x), \forall x \in X.$

The *extension principle* introduced by Zadeh (1965) is a general method for extending non-fuzzy mathematical concepts to the fuzzy framework.

Definition A2.12 (extension principle): Let $f : X \to Y$ be a mapping from a set *X* to a set *Y*. The extension principle then allows us to define the fuzzy set *B* in *Y* induced by the fuzzy set *A* in *X* through *f* as follows:

$B = \{(y, \mu_B(y)) \mid y = f(x), x \in X\},$

with

$$\mu_B(y) = \mu_{f(A)}(y) = \begin{cases} \sup_{y=f(x)} \mu_A(x) & f^{-1}(y) \neq \varnothing \\ 0 & f^{-1}(y) = \varnothing \end{cases},$$

where $f^{-1}(y)$ is the inverse image of *y*.

The concept of a *Cartesian product of fuzzy sets* can be introduced as follows:

Definition A2.13 (Cartesian product of fuzzy sets): Let $A_1,...,A_n$ be fuzzy sets in $X_1,...,X_n$ with the corresponding membership functions $\mu_{A_1}(x_1),...,\mu_{A_n}(x_n)$, respectively. The Cartesian product of the fuzzy sets $A_1,...,A_n$, denoted by $A_1 \times \cdots \times A_n$, is then defined as a fuzzy set in $X_1 \times \cdots \times X_n$ whose membership function is expressed by $\mu_{A_1 \times \cdots \times A_n}(x_{1,...,}x_n) = \min\left(\mu_{A_1}(x_1),...,\mu_{A_n}(x_n)\right).$