AN ALGORITHMIC APPROACH TO CRUSTAL DEFORMATION ANALYSIS

Huseyin Baki Iz
Department of Geodetic Science and Surveying

AN ALGORITHMIC APPROACH TO CRUSTAL DEFORMATION ANALYSIS

by

Huseyin Baki Iz

ABSTRACT

In recent years the analysis of crustal deformation measurements has become important as a result of current improvements in geodetic methods and an increasing amount of theoretical and observational data provided by several earth sciences.

In this study, a "first-generation" data analysis algorithm which combines a priori information with current geodetic measurements was proposed. Relevant methods which can be used in the algorithm have been discussed. Prior information is the unifying feature of this algorithm. Some of the problems which may arise through the use of a priori information in the analysis have been indicated and preventive measures were demonstrated.

The first step in the algorithm is the optimal design of deformation networks. As an example of deformation model oriented designs, it was shown that regular polygonal deformation networks composed of equilateral triangles are uniformly D-optimal for a homogeneous deformation field.

The second step in the algorithm identifies the descriptive model of the deformation field. A method based on the entropy measure of information, proposed by Box and Hill (1967), was applied to a group of postulated deformation models and the identification of the correct model was demonstrated through an example.

The final step in the algorithm is the improved estimation of deformation parameters. Although deformation parameters are estimated in the process of model discrimination, they can further be improved by the use of a priori information about them. According to the proposed algorithm this information must first be tested against the estimates calculated using the sample data only. Null-hypothesis testing procedures were developed for this purpose.

Six different estimators which employ a priori information were examined. Emphasis was put on the case when the prior information is wrong and analytical expressions for possible improvements under incompatible prior information were derived.

ACKNOWEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1
INTRODUCTION

1.1 BACKGROUND

Almost half a century after the theory of continental drift was first proposed by Wegener in 1912, the plate tectonic hypothesis was put forward by Harris Hiss, J. T. Wilson and others (NASA, 1978). Since then it has grown up mainly as a result of an overwhelming body of new geological and geophysical evidence that supports the plate tectonics idea.

The plate tectonics model depicts the outer shell of the earth, the lithosphere, as broken into small number of large plates moving relative to each other and with boundaries marked by the earthquake zones.

They converge along the seismically active continental margins and arcs. They diverge along the axes of the ocean ridges. They slide along each other in areas like the San Andreas rift zone in California (Drake, 1983).

Although the plate tectonics idea is based on a simple model, its surface manifestations as crustal deformations are complex in nature. In spatial spectrum, tectonic plates are assumed to behave as rigid blocks to the first order (global scale phenomenon). Their motions, as rigid entities, may reach 1-10 centimeters per year. Yet, they undergo severe deformations at their boundaries. Regional scale phenomena take place over distances less than the dimensions of typical plates, a few thousand kilometers, but larger than a few hundred kilometers. Local scale phenomena occur in the immediate vicinity of a fault and it is closely related to the regional deformations.

In the temporal spectrum, long term strain accumulations (1-10 years) are identified in regional and local scales. Medium term episodic

1

changes (less than a year) and short term precursory crustal motions before large earthquakes are possible (NRC, 1981).

In recent years, the improvements in geodetic instrumentation and measurement techniques have made feasible the detection of crustal movements over reasonably short time intervals. Hence, the interpretation of recent crustal movements is now in the realm of Geodesy.

Today the classical ground based geodetic methods provide relative position and position changes over short distances with a precision of 0.1 ppm (Slater et al., 1983). They are useful for determining regional and local strain rates on the plate boundaries. Space techniques are far superior than the classical techniques for determining plate tectonic motion and large scale deformation of the plates. They are precise enough (on the order of 0.1 ppm - 0.7 ppm for 200 - 6000 km baselines) to detect possible plate motions which may reach a few centimeters per year (Coates et al., 1985, Davidson et al., 1985).

Geodetic observations can provide information about the overall rate and direction of relative motions among the plates. It is now possible to investigate whether these motions vary along the length of the boundary, whether all these motions are accommodated or do all of the accommodations to plate motion take place along plate boundaries or are some distributed over a broad area. Detailed configuration of the strain field around the ends of locked segments of the fault zones can be mapped using geodetic observations. Temporal variations can be detected. Precursory strain changes in local and regional scales are important for understanding earthquake mechanisms.

Today there is an increasing number of repeated geodetic observations and additional information about the nature of crustal movements provided by several other earth sciences. The analysis and the interpretation of this accumulated data are now more important than ever. Meanwhile, a need for improvement of the current analysis techniques has already been recognized (NRC, 1981).

1.2 PURPOSE

The purpose of this study is to exploit the use of prior information in the analysis and interpretation of recent crustal movement measurements by the geodetic methods.

Two specific objectives are under consideration. The first is to develop a general data analysis algorithm in which prior information is introduced into the analysis of geodetic data. The second is to propose and elaborate on relevant methods which can be employed in this algorithm.

Most of the time geodesists possess some knowledge about the crustal deformation phenomenon under consideration. This knowledge may come from geological or geophysical investigations, from purely theoretical considerations or it may have been derived from other geodetical measurements. This type of knowledge is referred to as prior, auxiliary or extraneous information. It comes from outside the current geodetic observations themselves. It may be qualitative or quantitative.

The desirability of exploiting this extraneous information should be clear. The more the geodesists know, the more effectively they can plan and analyze. It is intuitive that a gain in efficiency would result at different stages of the investigation by the use of additional information. The introduction of quantitative information can improve the estimation of relevant deformation parameters under controlled circumstances.

Yet, the introduction of prior information needs serious attention since there is always a likelihood that it may not be compatible with the sample data, i.e., observations. The following questions may arise upon introduction of extraneous information:

1. If a variation of descriptive models is available in modeling crustal deformations, how should geodetic surveys be designed in order to discriminate among these models?
2. If prior information exists about the deformations, when should they be introduced into the analysis?

3

3. If prior information is chosen to be used in the estimation, which estimator should be used? What can be gained by using prior information?

4. If discrepancies exist between prior and the information implied by the geodetic measurements, can they still be combined to improve the analysis? If so, under what conditions?

Relevant methods to deal with the aforementioned questions are proposed. Their interactions are demonstrated in the proposed general data analysis algorithm.

1.3 SCOPE AND ORGANIZATION

Chapter 2 starts with a short review of the geodetic aspects of local, regional and global deformations. A general data analysis algorithm is then proposed. The overall approach of the algorithm is an inductive one. "Every solution of a problem raises new unsolved problems; the more so the deeper the original problem and the bolder its solution" (Popper, 1972). It is recognized that there are no complete solutions. Therefore the algorithm tests different models until one model is tentatively established.

The algorithm is composed of four major steps: design of geodetic deformation networks, discriminatory analysis of different models, diagnostic checking and improved estimation. The use of additional knowledge (qualitative or quantitative) is the unifying feature of this algorithm. The following chapters elaborate on these topics.

In Chapter 3, a D-optimal design for homogeneous deformation field is first derived. Then an entropy measure of information is used in a Bayesian setting which employs prior densities about the relevant deformation parameters as additional information for the sequential design of geodetic observations and for model discrimination purposes. The procedure is demonstrated through an example.

Chapter 4 discusses the use of quantitative prior information in the estimation for the purpose of improving the deformation parameters.

First, a statistical testing method, which checks the agreement between prior information about model parameters and their estimates from the current observations is given. Second, a comparison criterion is set up to express the preferences about the estimators. The problems which may arise due to the use of prior information, particularly bias in prior information, are discussed. Several new theorems are derived for possible improvements in such cases.

Chapter 5 considers the estimation problem when parameters are assumed to be stochastic in nature. A similar discussion follows as in Chapter 4.

5

Chapter 2
CRUSTAL DEFORMATIONS AND AN ALGORITHMIC
APPROACH TO THEIR ANALYSIS

2.1 LOCAL, REGIONAL AND GLOBAL DEFORMATION ANALYSIS

In this section the geodetic aspects of crustal movements are briefly reviewed. An ideal approach to the analysis of the kinematics of tectonic motions would include both horizontal and vertical components for more meaningful inferences. However, the horizontal aspect of crustal movements is the main theme in this study. This is mainly due to the fact that the plate tectonic hypothesis, which explains tectonic motions, is constrained to the surface of the earth. Vertical motions are weakly explained through the horizontal motions which are sometimes considered to be a failure of the plate tectonic hypothesis in part (Beloussov, 1979). However, the proposed approaches in this study are general enough to include any type of data if they are proven to be relevant.

First the existing approaches to the local, regional and global deformation analysis are briefly discussed. Then a general data analysis algorithm is proposed.

The uniformity of relative plate motions and the rigidity of the major plates are the two crucial unifying principles of global tectonics. However, these postulates are undoubtedly valid in the long-term average sense and for the average totality of the plates. Significant departures and elastic deformations in both regional and local scales are likely to occur. In fact, such elastic deformations and strain accumulations have already been confirmed by the interseismic deformation observations along plate boundaries using geodetic methods (NRC, 1981).

6

The changes of shape and dimensions of the geodetic networks are considered to be deformations of the terrain surface. These deformations may be uniform or vary with time. Short- and long-term monitoring of these networks at seismically active areas provide information about the nature of deformations.

Two different types of deformation networks are identified: relative networks in which all network points are located on the deformable object and fixed networks in which there are stable points outside the deformable object.

The interpretation of observed changes in geodetic observables in terms of crustal deformations are by no means unique. Today several approaches exist for the analysis of geodetic entities. The differences among these approaches are a result of the type of the network under consideration.

Let the position of the network points be given at an initial state by their projections (x_0, y_0) on the axes of a Cartesian coordinate system XY. Furthermore, let the network points along the same axes acquire the displacements u, v which are functions of the coordinates x_0, y_0 at every instant of time. The position (x, y) of an arbitrary network point at an epoch t is then determined in the same Cartesian system by the following coordinates

$$x = x_0 + u(x_0, y_0, t)$$

$$y = y_0 + v(x_0, y_0, t) \tag{1}$$

Displacement components u and v can be obtained by the difference of station coordinates adjusted at different epochs from the geodetic observations. Since physical realization of a coordinate system (fixed networks) on a deformable body is often very difficult, coordinate systems employed in these solutions should be carefully interpreted. Care should be given to use the same coordinate system for different adjustments regardless of whether or not they are realized as a result of minimum constraint, inner constraint or pseudo-inverse solutions.

A simple but not widely used approach in detecting network deformations is the conditional adjustment of observed quantities for each epoch. Their differences as a result of deformations may then be expressed in an appropriate coordinate system to display the relative displacement components of individual network points. Alternatively, a mathematical model can be set up in which displacement components appear as parameters. For instance, if baselines are observed at two different epochs, then the changes in these measurements may be expressed in terms of the displacement function as

$$d\ell_{ij} = \sin \alpha_{ij} \, dx_j + \cos \alpha_{ij} \, dy_j - \sin \alpha_{ij} \, dx_i - \cos \alpha_{ij} \, dy_i \qquad (2)$$

where $d\ell_{ij} = \ell_{i_0 j_0} - \ell_{ij}$. $\ell_{i_0 j_0}$ and ℓ_{ij} are the baseline measurements between two network points at epochs t_0 and t respectively. α is the azimuth of a baseline and

$$dx_j = x_j - x_{j_0}$$
$$dy_j = y_j - y_{j_0} \qquad (3)$$

If there are network points, either on the deforming surface or outside of it, which are observationally as well as statistically justified to be stable (Koch and Fritsch, 1980), then a reference coordinate system may be realized through these points. Otherwise, the displacement components are not estimable quantities since they transform under changes of coordinate systems (S-Transformations, Molenaar, 1981). Therefore, some type of minimum norm or minimum constraints solution is required. Deformations, on the other hand, are independent of the coordinate system in the sense of changes in size and shape.

Displacement components obtained through these procedures are ambiguous in the sense that they do not reveal much information about the nature of deformations. More insight is gained if the displacement field of the network points is approximated by means of algebraic polynomials (Ney, 1978).

Let the displacement function for the horizontal network points be written, for instance, for a polynomial of a second degree as

$$u(x,y) = a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2$$

$$v(x,y) = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 \tag{4}$$

where a and b are the coefficients of the polynomial function. It is possible to determine these coefficients from the individual network point displacements which are obtained through one of the previously outlined methods.

Alternatively, considering (1) and (3) and substituting (4) in (2), a new mathematical model is obtained in which coefficients of the approximating polynomial appear as parameters. A least squares solution of the resulting observation equations in a fixed network system with a sufficient number of observations yields to the determination of unknown coefficients. The question of the best fitting polynomial may then be answered by a variance test of the significance of estimated coefficients (Uotila, 1980).

Once the approximating polynomial is determined, it can then be used for interpolation purposes, for filling the observational gaps (Ellmer and Welsh, 1982), or for further inferences about the nature of tectonic motions. It may also be useful in determining the local interpolating polynomials of the finite element method.

Observed relative displacement of network points due to the tectonic motions are always small (except for sudden ruptures along the fault zones which occur during earthquakes). Furthermore, deformations which the earth's crust undergoes as a result of applied stresses can be considered to be elastic to a certain extent. Consequently, linear elasticity theory, which deals with the description of homogeneous strain field, becomes one of the basic tools in analyzing tectonic strain field and tectonic motions.

The interpretation of the observed changes in geodetic observables in terms of the crustal deformations is performed in this case by describing them through the horizontal strains. Although this approach, like the others, is not sufficient for coping with all the problems arising

in the analysis of tectonic motions, it is nevertheless better suited for further inferences since it provides spatial information in simple geometries and achieves a better presentation of results.

Let the displacement function be represented by the following polynomial

$$u(x,y) = a_0 + a_1 x + a_2 y$$
$$v(x,y) = b_0 + b_1 x + b_2 y \tag{5}$$

Since infinitesimal strains are, by definition, the linear functions of displacement gradients and are given by

$$e_x := \frac{\partial u}{\partial x} , \quad e_y := \frac{\partial v}{\partial y} , \quad e_{xy} := \frac{1}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right), \quad \omega := \frac{1}{2}\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) \tag{6}$$

substitution of (5) into (6) leads to the homogeneous deformation model which can be expressed as

$$u(x,y) = a_0 + e_x x + e_{xy} y - \omega y$$
$$v(x,y) = b_0 + e_{xy} x + e_y y + \omega x \tag{7}$$

In this representation, the parameters have the following geometrical meanings: a_0 and b_0 correspond to translation elements, e_x and e_y are the extensional components of infinitesimal strain along the X and Y axes respectively and are positive for extension and e_{xy} is the shearing strain, which is positive for the right lateral shear. ω represents the infinitesimal rotation of the network points (in an average sense). When the above quantities refer to a certain period of time with uniform rates, the symbols are marked with a dot. The dimension of all strains is micro strain, or part per million, or micro strain per year. There is also a multitude of other representations of deformations, such as engineering shear, principal strains, etc. (see Pope, 1966 and Welsch, 1982 for a detailed description of homogeneous horizontal infinitesimal strains). These elements in the sequel will be called deformation parameters.

Deformation parameters can be calculated by first determining the displacement components and then solving for the strain elements or for some other representations using these network point displacement components as pseudo-observations. Alternatively, they can be obtained directly from a mathematical model which is obtained, for instance, by the substitution of (7) into (2). Both approaches yield identical results. Their differences are due to manipulative convenience (Brunner et al., 1980). This is true, however, only as long as the same estimation techniques are used in both approaches. As we shall see later, there are alternative techniques besides least squares. Results are likely to be different if different estimation techniques are used at different stages. In such conflicting circumstances it is intuitively clear that the direct estimation of the relevant deformation parameters is preferred since displacement components play a transient role in the analysis.

The estimability of the relevant deformation parameters in this representation are again determined by the type of geodetic network and the type of geodetic observations. If, for instance, baselines are observed in a relative network and furthermore their scales are compatible at different epochs through calibration procedures, then the principal strains, total shear and dilatation are invariant quantities under coordinate transformations. They are therefore estimable. For a complete discussion of estimable quantities in infinitesimal strain field confer: Livieratos (1980), Dermanis (1981) and Welsch (1982).

In any event, direct representation of repeated crustal movement observations in terms of estimable deformation parameters (the ones which are invariant under S-Transformations) is the simplest approach to avoid ambiguities which may arise due to the difficulties in realizing a coordinate system for the relative networks. Nevertheless the estimation of coordinate system dependent deformation parameters may sometimes be desirable (such as extensional infinitesimal strain normal to the fault trace). In this case a suitable coordinate system is introduced accordingly at a fundamental epoch. As a result, it is only natural to use at any time instant of observation the same coordinate system. Otherwise, these parameters are not comparable from one epoch to the

11

other. In this framework, the coordinate system introduced becomes an integral part of the descriptive model of deformations under consideration.

The general procedure for calculating strain or displacement field on a regional scale can be achieved using the above approaches in 3-D (Harvey, 1985). However, treating the network as a whole as homogeneous may not be realistic in areas where complex faulting systems exist. In this case a nonhomogeneous (homogeneous with discontinuities) strain field is modeled that best accounts for the observed changes in observations (Chrzanowski et al., 1983). Also, the treatment of the strain field as homogeneous over the whole area covered by the network implies averaging relevant deformation parameters. If such assumptions are considered to be unrealistic, the area of interest can be dissected into smaller elements, such as triangles. These are then analyzed either individually in local scale, or a finite element solution which involves the geometric aspect of displacement field is possible (Welsch, 1983).

In the case of global deformation analysis, it is assumed that the lithospheric plates move as rigid bodies on the asthenosphere with respect to each other. The rigid body motion of these tectonic plates, as a first approximation of global crustal movements, is the crucial assumption of the plate tectonics idea. After all, the entire concept of plate tectonics would not have much meaning if there are points which move relative to each other with velocities comparable to the velocities of relative plate motions. Therefore the axiomatic modeling of crustal motions on a global scale is much simpler than the local and regional scale phenomena.

Since long baselines can today be measured with sufficient accuracy using space geodesy techniques, the plate tectonics theory becomes the initial working hypotheses in the analysis of repeatedly measured long baselines.

Let $\bar{r}_i$ and $\bar{r}_j$ be the position vectors of observing stations on different tectonic plates expressed in an adopted coordinate system XYZ. Then the chord length for a baseline vector $\bar{r}_{ij}$, at an epoch t,

12

connecting two different plates is given by

$$r_{ij}^2 = (\bar{r}_i - \bar{r}_j) \cdot (\bar{r}_i - \bar{r}_j) \tag{8}$$

Linearization of this model about the baseline vector $\bar{r}_{i_o j_o}$ referring to an initial epoch t_o yields

$$r_{ij} - r_{i_o j_o} = \frac{\bar{r}_{i_o j_o}}{r_{i_o j_o}} \cdot d\bar{r}_i - \frac{\bar{r}_{i_o j_o}}{r_{i_o j_o}} \cdot d\bar{r}_j \tag{9}$$

where $d\bar{r}_i = \bar{r}_i - \bar{r}_{i_o}$, $d\bar{r}_j = \bar{r}_j - \bar{r}_{j_o}$ are the infinitesimal displacement vectors as a result of global tectonic motions. r_{ij} and $r_{i_o j_o}$ are the observed baseline lengths at epochs t and t_o respectively. Now if the observational errors u in baseline measurement differences y are considered, then the above model (9) leads to the general Gauss–Markov setup of linearized observation equations in which components of station displacements appear as parameters x to be estimated. For each baseline observed at epochs t and t_o, one observation equation is formed,

$$y = Ax + u \tag{10}$$

The solution of this system, in a least squares sense, can be achieved by fixing at least one plate and having at least two stations on each plate (three repeated baseline observations). An inherent assumption in the above model is that the scale of the baseline measurements is the same at both epochs. This can be achieved by a priori field calibration procedures. Parameters to be solved in this case are the relative displacement vector components referenced to the fixed plate.

Alternatively, minimum norm solutions are possible. Through this method the coordinate system is defined in an optimal fitting sense. Similarly, if additional information exists about the displacement vector for at least one plate a unique solution for the above set up is again possible (Bock, 1982).

Nevertheless, as in the case of local and regional deformation analysis, recovered displacement components are ambiguous as they

13

stand. They do not reveal much information about the nature of plate motions. They simply supply refined quasi-observational data though which kinematical or theoretical models can be tested. However, if such models are postulated beforehand, they can then be expressed directly as a function of baseline observation differences.

Of such models, instantaneous plate kinematics is the most popular in today's studies. In this approach, displacements assumed to be constrained on a spherical earth model are the results of rigid motion of plates. Consequently, the motion of the plates can be described by an axial rotation following Euler's theorem (Goldstein, 1965).

Consider $d\bar{r}_i$ and $d\bar{r}_j$ to be changes in the position vectors for stations located on different plates i and j respectively. Then the following relationships hold

$$\frac{d\bar{r}_i}{dt} = \bar{r}_i \times \bar{\Omega}_i$$

$$\frac{d\bar{r}_j}{dt} = \bar{r}_j \times \bar{\Omega}_j \tag{11}$$

where $\bar{\Omega}_i$ and $\bar{\Omega}_j$ are the pseudo-vectors of angular velocities on plates i and j respectively. Their Cartesian components are given by

$$\bar{\Omega} = \begin{bmatrix} \Omega_1 \\ \Omega_2 \\ \Omega_3 \end{bmatrix} = \begin{bmatrix} \dot{\omega} \cos \phi \cos \lambda \\ \dot{\omega} \cos \phi \sin \lambda \\ \dot{\omega} \sin \phi \end{bmatrix} \tag{12}$$

ϕ and λ are the spherical coordinates of instantaneous rotation axis and $\dot{\omega}$ is the magnitude of the angular pseudo-vector.

Substituting (11) into (9) with the assumption of linear velocities results, after some simple manipulations, in the following mathematical model of instantaneous plate kinematics

$$r_{ij} - r_{i_o j_o} = \frac{\Delta t}{r_{i_o j_o}} \left[\left(\bar{r}_{i_o} \times \bar{r}_{j_o} \right)' \vdots \left(\bar{r}_{i_o} \times \bar{r}_{j_o} \right)' \right] \begin{bmatrix} \bar{\Omega}_i \\ \bar{\Omega}_j \end{bmatrix} \tag{13}$$

where $\Delta t = t - t_0$. In this expression components of angular velocity pseudo-vector appear as parameters to be estimated as a result of the hypothesis which is put forth.

Estimation of these parameters follows the same arguments as in the case of estimation of displacements components. (13) can also be expressed as a function of spherical coordinates using the formulas of geometric geodesy (Drewes, 1982).

So far, current approaches to the representation of tectonic motions have been briefly discussed within the geodetic framework. Common to all these methods is the assumption of uniform time dependency. It is, however, recognized that strain accumulations at local and regional scales are likely to be nonlinear in time (Thatcher, 1983). There is also the possibility of nontectonic motions of network points at all scales. Therefore, further refinements in such models become a necessity.

These problems, along with some new methods in the analysis of tectonic motions, are discussed in the following proposed general deformation analysis algorithm.

2.2 AN ALGORITHMIC APPROACH TO CRUSTAL DEFORMATION ANALYSIS

It is widely recognized today that most of the information about the nature of crustal movements is supplied by the repeated geodetic surveys (NRC, 1982). Providing a choice of reliable quantitative data about the network deformations has been the traditional undertaking of geodesy in crustal dynamics studies. Beyond this routine task is that of a growing necessity to conjure up more detailed operational hypotheses about the kinematics of tectonic motions by geodesists. In the meantime, increasing information provided by other disciplines and different techniques can now be introduced into the analysis of current geodetic observations for more meaningful inferences.

In the present chapter, an algorithm which combines additional information and sample data is proposed. It is a more general approach in comparison of the current trends in analyzing geodetic data.

15

The algorithm is composed of the following stages (see Figure 1): deformation model oriented network design, discriminatory analysis of several concurring models, diagnostic checking and improved estimation. Relevant methods that can be used in this algorithm are exploited in the following chapters.

The identification of a suitable model of the deformation process which fits well to the changes in the observed geodetic quantities and also explains the differences between local regional and global scale components of tectonic motions is the main purpose of crustal movement analysis. Predicted results derived from such models can then be compared with theoretical data supplied by other disciplines and thus be used to test the validity and credibility of the fundamental assumptions of geotectonics theory.

Since the underlying physical phenomenon which causes the plate motions and the deformations along plate boundaries is still poorly understood, the modeling efforts in all scales are to be descriptive. These descriptive models, if not capable of constructing the actual phenomena by the critique of repeated geodetic surveys, do at least play an essential role in demonstrating what hypothesis should not be put forth. Questions such as: "Does the rate of tectonic motions observed change with time?", "Does the rate of movements vary from one place to the other?" and "What can be learned from them about the nature of driving forces?" lend themselves to different models.

Sometimes discrepancies found in different scales can be due to the coarseness of the model used. In such cases more detailed models are needed to explore the possible modes of deformations, especially in the active plate margins where severe deformations are likely to occur.

Other times there may be more than one descriptive model which concurs with the differences of repeated geodetic surveys. Thus, the modeling efforts are unescapably iterative in nature on different postulates.

Although postulating a general class of models is mostly based on the experience and intuition of an experimenter, the qualitative information provided by other disciplines and previous in situ

16

measurements play an important role in the identification of different models. Discrepancies found in the interpretation of different types of information can help geodesists to postulate such models.

The selection of the best model can be achieved by performing repeated geodetic observations and using statistical tests on estimated parameters of different models. Therefore, the model discrimination process is intimately related to the design of optimal deformation networks and optimal observations.

Model oriented deformation network designs are necessary in this respect. However, such designs are still not differentiated from the classical network designs, mostly due to the fact that deformation parameters are obtained from the adjusted coordinates in a two step procedure as summarized in the previous section. A model oriented D-optimal design for the homogeneous deformation field is constructed analytically in Section 3.1 as an example.

Meanwhile, such designs may not be satisfactory if they are confined to a single model of the displacement field. Therefore, it is also necessary to develop designs which do well for different competing models. However, analytic construction of such designs is difficult even in simple cases. Iterative methods for constructing model robust linear optimal designs are given in a series of articles by Cook and Nachtsheim (1982) and Lauter (1974).

The next step in the algorithm is the identification of the best working hypothesis among the others. Since these are temporal models, their identification should be a result of sequential experimental designs. That is, the observations and the analysis are designed in stages. In this way, the data collection procedure can be carried out in an optimal fashion since the geodesists set forth the end-purpose of the investigation being conducted.

First, models of the deformation process are postulated. A network design which does well for such models is set up. This initial design is surveyed at least at two different epochs and the models are evaluated. The tectonic motions observed in nature are always small and very often they remain well below the observational accuracies of geodetic methods.

17

Furthermore, they are likely to be temporally nonlinear, such as in exponential decay in strain rates (Thatcher, 1983). Therefore preference of one model over the others is not conclusive if it is based solely on observations performed at a few epochs. The data may tend to favor one model over the others as the number of surveys increase. This information can be used to design new observations. This model discrimination procedure is carried out until one model is selected. Each time new observations are designed and performed using the information gained from the previous resurveys and model evaluations. Such a method of model discrimination is formulated in Section 3.3 in a Bayesian setting using the entropy concept of information.

Descriptive modeling is quantified by the estimation of the relevant parameters. These two processes are mutually exclusive. Once the parameters are estimated, they can in turn be used to design new improved models and better estimates. The limiting factor in this process, first lies in the accuracy of geodetic measurements. Their uncertainties include contributions from both systematic and random errors which can be reduced effectively only up to a certain level (Mierlo, 1975; Baarda, 1975). Secondly, the coverage of deformation measurements for monitoring tectonic motions is still confined at particular regions because these techniques are still time consuming and expensive. Therefore, such a finite number of samplings may not reveal the overall behavior of the deformations since they are representative of only a regional and temporal average sense.

On the other hand, deformation parameters estimated from these sample observations can further be improved to some extent by the introduction of extraneous information in the estimation (it is obvious that there is no need to introduce additional information if the model parameters are found to be satisfactory). Perhaps the most important example of the use of prior information in geodesy may be proven to be in crustal dynamics since the process includes many entities of the lithosphere in a complex way, and it involves several disciplines. Therefore, there is an abundance of qualitative and quantitative

18

information in addition to the information provided by the geodetic methods.

The qualitative aspect of this information is useful, as discussed, in postulating different models. Quantitative information may arise in various ways in crustal movement analysis. Sometimes theoretical solutions such as global motion of plates based on the driving force models for plate tectonics (Solomon and Sleep, 1974; Kaula, 1975) or the global intraplate stress fields predicted by various force systems, other times the results of past samples such as in situ measurements of strain through extensometers, resurvey of old geodetic networks or examination of stress induced geological structures, are thought to give information about parameters. In all these cases, prior information does not arise from the particular body of data currently being analyzed. If this information is correct, then it is surely useful to incorporate it into the statistical estimation procedures as well as physical modeling of such phenomena. Such prior information may be valuable in increasing the precision and the reliability of the estimates, particularly when samples are limited in size.

Therefore, the study of valid prior information becomes an integral part of the proposed algorithm which will be referred to as diagnostic checking in this study. In this step of the algorithm, quantitative prior information should be identified and tested against the sample information. A null-hypothesis testing procedure is given in Section 4.2. If prior and sample information are found to be compatible then they can be combined in the improved estimation of parameters.

However, noncompatibility as a result of such tests by no means implies that prior information is erroneous and should be totally disregarded. Testing procedures themselves are likely to contain further uncertainties, after all. For example, probability distributions attributed to the sample observations may not be realistic or significance levels, if used, may be too small compared to the signal to noise ratio. There may also exist the possibility that the information contained in current data is different than the additional information simply because the latter is representative of tectonic motions only in

19

the long-term average sense. Similarly, prior information as a result of global solutions may be different than the small scale information because of the local properties of tectonic motions.

Consequently, prior information in practice is not known with certainty. What then is the purpose prior information if we can never be sure of the validity of such constraints? The answer lies in the fact that additional data may always contain a certain component of reality. If the value of it is found to be small then it is clearly unwise to use dubious information. However, if the value of improvement is large, then it may pay to use even dubious information.

A better alternative is to use estimation techniques which are insensitive to wrong prior information about parameters but which still result in better estimates than those using sample observations only.

At this point, it is necessary to identify two different concepts in estimation. First, the well known approach, namely, estimation when the model parameters are deterministic; and second, estimation when the model parameters are considered to be stochastic (random). In the latter case, statistical model setups are known to be random effect models (Searle, 1974; Harville, 1976). Estimation of random parameters is also called prediction (Searle, 1974; Schaffrin, 1983).

So far implicit in the discussion is that the mechanism of the underlying phenomena is deterministic but unknown to the user. The representation of displacement field with stochastic parameters does not contradict with this view.

> *There is no conflict between causality or randomness or between determinism and probability if we agree, as we must, that scientific discoveries are not discoveries of nature but rather inventions of human mind. Their consequences are presented in deterministic form if we examine the results of a single trial; they are presented as probabilistic statements if we are interested in averages of many trials...* (Papoulis, 1984).

Similarly, in crustal movement analysis, averaging strains over whole areas as well as over certain periods of time can be interpreted to mean

strain parameters are random with certain mean and variances. This becomes much more clear in the case of global plate kinematics. In this process the deformations in temporal and spatial spectrums can be interpreted as smooth motions as trends over geological time scales with superimposed short term jerky motions (Kaula, 1978). Solutions such as numerical analysis of instantaneous plate kinematics (Minster and Jordan, 1978) can be recognized as the expected values of possible tectonic motions whereas current global observations are contaminated, in addition, with short term random effects.

Therefore, information as a result of temporal and spatial averaging can be introduced into the analysis of current tectonic motions within this interpretation. However, in all likelihood such information may be wrong. Hence, it is desirable to have estimates which are not dominated by erroneous prior information.

Of course estimators with such virtues are hard to find. Moreover, it would be a mistake if one estimator or predictor is singled out among the others and used under all varying conditions. Therefore the study of suitable estimators, their behavior under different types of prior information and the gain from using prior information becomes the last part of the proposed algorithm. These problems are examined in Chapters 4 and 5.
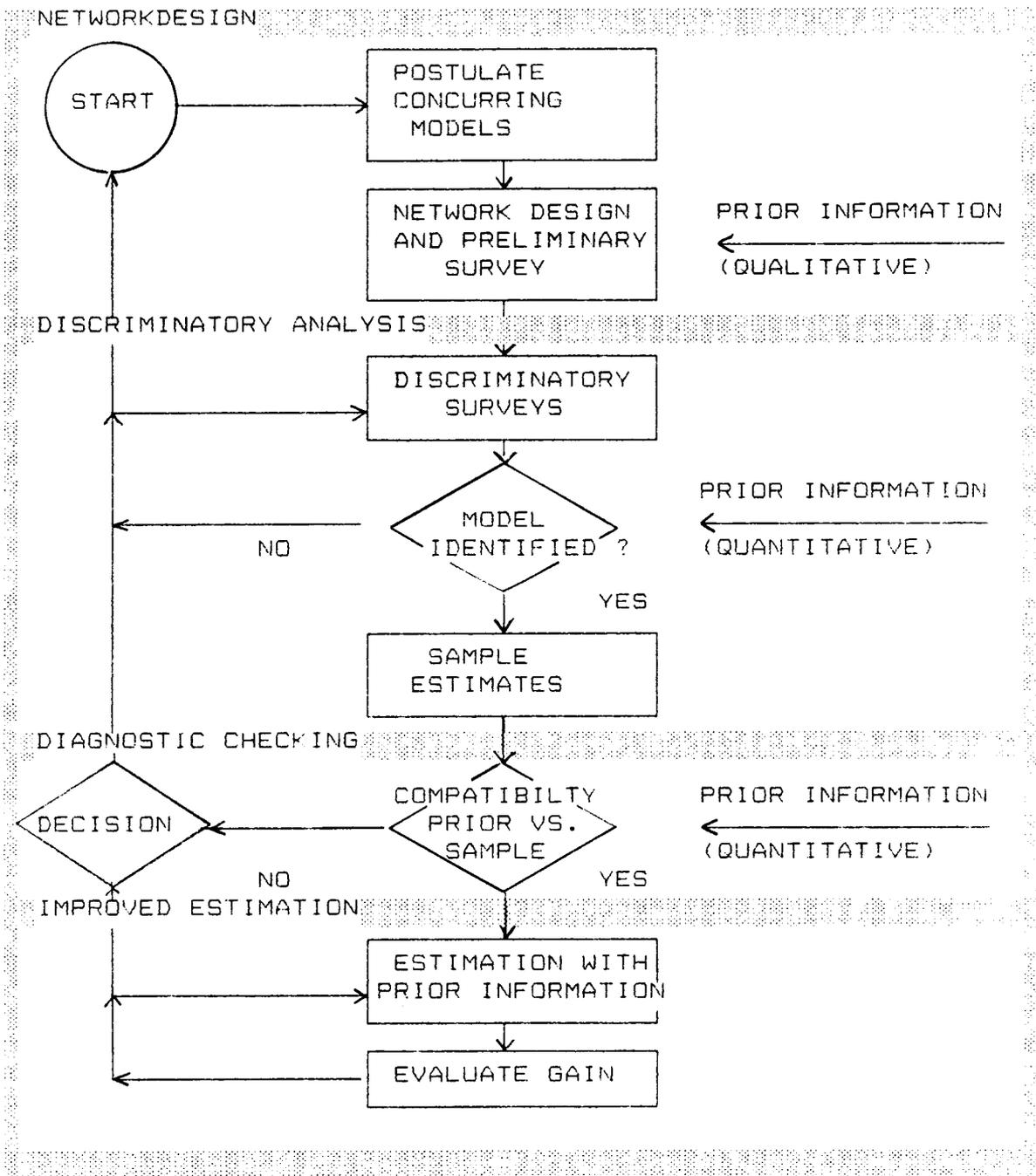
Figure 1. Algorithm for Crustal Movement Analysis

22

Chapter 3
DESIGN AND MODEL DISCRIMINATION

3.1 D–OPTIMAL DESIGNS FOR HOMOGENEOUS STRAIN FIELD

A geodetic network is defined through geodetic measurements between observation points and the definition of a datum (coordinate system). Geodetic networks are designed according to a particular task under consideration, type of measurements and their accuracies using a predefined optimization criteria. Optimization of geodetic networks are well established in the geodetic literature (see (Grafarend et al., 1977) for a comprehensive review). Design of deformation networks can also be examined within the existing approaches. Bock (1982), using approximate theory for optimal designs, has established D and A optimal polyhedra designs for the distribution of crustal movement monitoring stations over the earth. However, as discussed in Chapter 2, the two-step procedures (first point displacements in terms of coordinate differences at different epochs are determined then the deformation model parameters are solved using the results of the first step as quasi observations) can equivalently be replaced by the direct estimation of relevant parameters. In this case it is natural that optimal design of deformation networks should be oriented to these particular models of deformation. But this problem has not been fully exploited in the geodetic area.

In this section, a model oriented D-optimal network design is constructed as an example of the homogeneous deformation field model using probability measures.

Consider the linear model

$$y = Ax + u , \quad u \sim (0, \sigma_u^2 I) \tag{1}$$

23

where y is an n×1 vector of observations, A is an n×m design matrix of full column rank and is composed of 1×m vectors of rows (a_i) depending on the form of the model assumed. x is an m×1 vector of unknown parameters. u is the vector of independently and identically distributed random variables with mean zero and variance σ_u^2. The experimental region is denoted by χ and a_i are continuous on χ. Then from the ordinary least squares, the dispersion matrix of $\hat{x}$ is

$$D(\hat{x}) = N^{-1} \tag{2}$$

where $N := \sigma_u^{-2}(A'A)$. In subsequent discussion N will be called the information matrix of the experiment e and will be denoted by N(e).

The design problem, therefore, consists of selecting vectors of control variables $a_i \, \varepsilon \, \chi$, i = 1, ..., n such that the design defined by these n vectors is in some defined sense optimal.

<u>Definition 1</u>: We define the optimal design for deformation networks as the one that minimizes the determinant of the covariance matrix of parameters.

Let n denote the number of repetitions of the ith observed quantity such that $N = \sum_{i=1}^{n} n_i$ and N is the total number of observations; then,

<u>Definition 2</u>: (Fedorov, 1972). A normalized design e is the collection of variables p_1, p_2,..., p_n, a_1, a_2, ..., a_n, where $\sum_{i=1}^{n} p_i = 1$ and $p_i = n_i/N$.

<u>Theorem 1</u>: (Fedorov, 1972). The family of matrices N(e) corresponding to all possible normalized designs, is convex.

<u>Corollary 1:</u> If $|N(e)|$ is the determinant of the design of an experiment, then maximizing $|N(e)|$ is equivalent to minimizing $|N(e)|^{-1}$.

<u>Definition 3</u>: (Silvey, 1980). The design that maximizes $|N(e)|$ is called D-optimal. If the errors are normally distributed then a confidence ellipsoid for the parameters x (of given confidence coefficient and for a given sum of squares arising from the design e), has the form

$$\{x: \quad (x - \hat{x})'N((x - \hat{x}) \leqslant \text{constant}\} \tag{3}$$

where $\hat{x}$ is the least squares estimate of x. The content of this ellipsoid is proportional to $|N|^{-\frac{1}{2}}$. Hence, D-optimal design makes this ellipsoid as small as possible. In summary, e is D-optimal iff $N(e) > 0$ (positive definite) and

$$\max_{a \, \varepsilon \, \chi} \quad |N(e)| = |N(e^*)| \tag{4}$$

where e^* is the optimal design. Since $u \sim (0, \sigma_u^2 I)$, D-Optimal design is also a minimax design (Kiefer, 1961).

<u>Definition 4</u>: (Fedorov, 1972). e is G-optimal iff

$$\min_{e} \quad \max \quad d(a,e) = \max_{a \, \varepsilon \, \chi} \quad d(a,e^*) \tag{5}$$

where $d(a,e)$ is the variance of the expected response $E(y)$. In words, the optimal design e minimizes maximum of $d(a,e)$. A sufficient condition for e to satisfy G-optimality is

$$\max_{a \, \varepsilon \, \chi} \quad d(a,e^*) = m \tag{6}$$

<u>Theorem 2</u>: The general equivalence theorem (Kiefer, 1961): Conditions 4, 5 and 6 are equivalent.

We are now ready to set up a semi-intuitive approach to determine an optimum design for the configuration of deformation networks with baseline observations satisfying D-optimality criteria. By consideration of continuous analogs of discrete designs we will arrive at certain conditions under which discrete designs can be constructed. Such an approach reflects the approximation of continuous designs with a distinct number of points in their spectrum.

<u>Definition 5</u>: A design measure is a probability measure denoted by ξ on χ. Specifically, $\xi \, \varepsilon \, \{D\}$ of all measures defined on the Borel field B,

25

generated by the open sets of χ and such that

$$\int_\chi d\xi(a) = 1 \qquad 0 < \xi < 1 \tag{7}$$

Given a measure in D the information for the experiment e will be taken as

$$N(e) = \int_\chi a'a \, d\xi(a) \tag{8}$$

When $\xi(a)$ is an absolutely continuous distribution, it has a probability density function given by $p(a)$. Equations (7) and (8) then reduce to

$$N(e) = \int_\chi a'a \, p(a) \, da$$

$$\int_\chi p(a) \, da = 1 \qquad p(a) > 0 \tag{9}$$

In the case of a design with finite spectrum having n points

$$N(e) = \sum_{i=1}^n p_i a_i' a_i \tag{10a}$$

$$\sum_{i=1}^n p_i = 1 \qquad 0 < p_i < 1 \tag{10b}$$

Then for any discrete design e it is possible to form a measure ξ by attaching a mass n_i/N to each point of e. From these definitions we see that if a continuous design is found then the discrete designs can be constructed by solving the equations in (10) for a and p.

Consider now the mathematical model for repeated distance observations at two different epochs (Pope, 1966)

$$d\ell = \ell \sin^2\alpha \, e_x + \ell \cos^2\alpha \, e_y + \ell \sin 2\alpha \, e_{xy} \tag{11}$$

26

This model appears as a trigonometric function in azimuths where e_x, e_y and e_{xy} are the deformation components of the infinitesimal strain tensor to be estimated and

$$a_i = \ell [\sin^2\alpha_i \quad \cos^2\alpha_i \quad \sin 2\alpha_i] \tag{12}$$

α_i is the azimuth of an arbitrary baseline observation ℓ and $d\ell = \ell - \ell_0$ is the difference of two baseline observations performed at two different epochs. Then the information matrix for these pseudo-observations is

$$N(e) = \int_X a'(\alpha) \, a(\alpha) \, d\xi(a) \tag{13}$$

If a uniform measure

$$d\xi(a) = (2\pi)^{-1} \, d\alpha \tag{14}$$

such that

$$\int_0^{2\pi} d\xi(a) = 1 \tag{15}$$

is selected in the domain of all possible azimuths $[0,2\pi]$, then from (13)

$$N(e) = (2\pi)^{-1} \int_0^{2\pi} a'(\alpha) \, a(\alpha) \, d\alpha \tag{16}$$

Now considering the following relationships

$$\int_0^{2\pi} \sin kx \cos \ell x \, dx = 0 \qquad \ell \neq k$$

$$\int_0^{2\pi} \sin^2 kx \, dx = \int_0^{2\pi} \cos^2 kx \, dx = 0 \tag{17}$$

it can easily be shown using (12) and (16) that

$$d(\alpha_i, e) = a_i(\alpha) \, N(e)^{-1} a_i'(\alpha) = m = 3 \tag{18}$$

27

Therefore, baseline observations with the above uniform measure are G-optimal. From (4) and the equivalence theorem it follows that uniform design is D-optimal and minimax.

Now a discrete design can be found. The azimuths α_i and point masses p_i must be such that the following equations hold (from equation 10 and 12)

$$\sum_{i=1}^{n} p_i a_{11} = \sum p_i a_{22} = \frac{3}{8} \sum \ell_i^2, \quad \sum p_i a_{12} = \frac{4}{8} \sum \ell_i$$

$$\sum p_i a_{13} = \sum p_i a_{23} = 0 \qquad \sum p_i = 1, \; i = 1, 2, \ldots, n \qquad (19)$$

It is not difficult to verify that for $\ell_1 = \ell_2 = \ell_3$, $n = 3$ and $p_i = 1/3$, the above equations are satisfied. It follows that an equilateral triangle is D-optimal. Similarly, regular polygonal designs composed of equilateral triangles are also uniformly D-optimal for the homogeneous strain field. It can be shown that certain arbitrary designs of equilateral triangles are D-optimal. But uniformity of the design is no longer valid for these configurations, i.e., $p_i \neq 1/n$.

Since the above approach is oriented to the optimal configuration of deformation networks in the homogeneous deformation field, it can be classified under the first-order design problem of geodesy. In the case of $p_i \neq 1/n$, resulting configurations composed of equilateral triangles can be classified within both first- and second-order design problems by interpreting p_i as observational weights. Meanwhile this approach is not sufficient if more than one model which can represent the deformation field, as recognized and discussed in the proposed algorithmic approach, are postulated. In this case, it is necessary to design deformation network oriented network designs which does well for different competing models. Such designs can be constructed iteratively using also prior preferences as weights on each model. This problem can also be examined, as proposed by Schaffrin (1986, private communication), within the reliability aspects of geodetic networks (Baarda, 1977). This problem will not be discussed in this study.

3.2 SEQUENTIAL MODEL DISCRIMINATION

In Chapter 2 the importance of model discrimination in crustal movement analysis was discussed in conjunction with the proposed algorithm. Currently five major approaches to the model discrimination problem can be identified (Chrzanowski, 1981):   Delft, Hannover, Karlsruhe, Munich and Fredericton.

The Delft approach (Mierlo, 1978; Kok 1977, 1980) consists of first testing each single network measurement by so-called data snooping; second, testing the stability of the reference points using data snooping again but this time treating coordinates from individual network adjustments as observations; third, testing of single point displacements at each point of the network; and fourth, testing of deformation models which are postulated a priori as systematic relative displacements of a group of points in the network during an epoch.

The Hannover approach (Pelzer, 1974; Niemeier, 1979) is based on global congruency tests using analysis of variance techniques. First an outlier check of observations for each single epoch is performed, and the observational weights within and between epochs are determined. Second, the global congruency of the network geometry between the epochs is tested to detect significantly moved network points at different epochs with respect to the mean geometry of the network. Then the corresponding displacements are computed.

The Karlsruhe approach (Heck, 1980) is based on the analysis of variance and use of confidence regions for point displacement vectors. The approach consists of the following steps. First, measurements for each epoch are adjusted for outlier control and for the final determination of observational weights.   In the second stage, the internal stability of a group of reference points is tested.   Then a simultaneous adjustment of all epochs is performed using the stable network points as reference points.   From the resulting coordinate differences of the other network points, displacement vectors with respect to these fixed points and their corresponding confidence regions are computed and displayed for the visual interpretation of the resulting deformations.

29

The Munich approach (Welsch, 1982) puts emphasis on the description of deformations in consideration of coordinate system and geodetic datum invariants. First the existence of significant deformations is checked using a datum invariant statistical test. Repeated observations are then modelled as a function of certain parameters of the deformation field which are again invariant with respect to the underlying geodetic datum and coordinate system. These parameters are then estimated for each epoch together with their significance based on a certain level of probability.

The Fredericton approach (Chrzanowski, 1981; Chrzanowski et al., 1983) combines some of the aspects of the previous approaches in a more general algorithm. In addition it analyses and interprets the resulting changes in the repeated observations within the infinitesimal homogeneous as well as nonhomogeneous deformation field. First, observational data is screened for outliers and possible trends of deformations are investigated. Second, single point displacements or relative displacements of the network points are determined employing the changes of the observed quantities at different epochs as observations using a minimum constraint solution. The resulting displacement components are then used as quasi observations which are then modelled as a result of postulated deformation models. Finally these deformation parameters are estimated for each model and the best fitting model is selected using the results of global statistical tests and the calculated significance levels of deformation parameters for each model.

In summary, the existing approaches to the analysis of crustal deformation measurements are oriented first to the detection of crustal movements and in a limited sense to the descriptive model discrimination problem. However, in the analysis of statistical decision theory, two problems are generally distinguished. One is the problem of making the best decision (identification of an appropriate model in our case) on the basis of a given set of data, and the other is the problem of designing the best experiments in order to get information upon which a decision will be made. In this respect model discrimination procedures in an

30

algorithmic approach should include these two problems. But the existing approaches to the analysis of crustal movements address only the problem of model discrimination. A combined analysis based on the design of measurements to facilitate the discrimination procedures within the scope of an experimental design concept has yet to be exploited.

In this section, a method based on Bayesian philosophy and entropy measure of information is proposed for the elucidation of time-dependent models of crustal motions. The method, which is due to Box and Hill (1967), combines model discrimination and design of measurement concepts in a sequential strategy. First the method is reproduced, then the strategy is adapted to the time dependent models of crustal movements through an example.

Assume that geodetic surveys are performed at different epochs. The data indicate that crustal motions occurred during these time intervals and a set of concurring descriptive models are postulated either as a result of the displayed dislocation patterns of the network points and/or previous information. The question is how to use the data to select the best model and to design new optimal observations to facilitate model discrimination.

Since in general every decision is a result of some type of decision rule, it is necessary at this point to define a discrimination criteria. Let the state of the deformations be represented by the parameter vector x and suppose that H is the set of concurring models (hypothesis). The preference of one model over the others (best model) can be ranked by means of a loss function $L(H,x)$. If

$$L(H_1,x_1) < L(H_2,x_2) \tag{20}$$

holds, then H_1 is preferred to H_2 where H_1, H_2 ε H. But since x is unknown and must be estimated from a set of observations y which are random variables due to the observational errors, it is intuitively much more appealing to make a decision based on the smallest possible expected loss over all conceivable set of samples S rather than loss only. Therefore,

31

$$r(L,\mathrm{x}) := \int_S L(\mathrm{H},\mathrm{x}) f(\mathrm{y}|\mathrm{x})\ d\mathrm{y} \tag{21}$$

where $r(L,\mathrm{x})$ is the expected loss (or risk) and $f(\mathrm{y}|\mathrm{x})$ is the density function of y given x. Thus, the problem of selecting the best model can be judged by examining the above risk function when x is unknown.

However, if some a priori information is already available about x before the observations are performed, this information can be used to make the best selection after the data is available. If, in particular, this information is a probability function on x then,

$$R(L,\mathrm{x}) := \int_P \int_S L(\mathrm{H},\mathrm{x}) f(\mathrm{y}|\mathrm{x}) f(\mathrm{x})\ d\mathrm{y}d\mathrm{x} \tag{22}$$

or considering (21)

$$R(L,\mathrm{x}) = \int_P r(L,\mathrm{x}) f(\mathrm{x})\ d\mathrm{x} \tag{23}$$

where $f(\mathrm{x})$ is a multivariate density on the parameter space P representing the prior information and $f(\mathrm{y}|\mathrm{x})$ is now interpreted as the conditional density of y given x. Again, it is possible to make a selection, this time in the presence of prior information, which minimizes the risk given in (23). This approach in general is known as the Bayesian solution. Equivalently, considering the Bayes formula, the conditional distribution of x given y is defined as

$$f(\mathrm{x}|\mathrm{y}) = \frac{f(\mathrm{y}|\mathrm{x})f(\mathrm{x})}{f(\mathrm{y})} \quad \text{if } f(\mathrm{y}) > 0 \tag{24}$$

where

$$f(\mathrm{y}) = \int_P f(\mathrm{y}|\mathrm{x})f(\mathrm{x})\ d\mathrm{x}$$

is the marginal density on y. Then (23) takes the following form,

$$R(L,\mathrm{x}) = \int_S \left[\int_P L(\mathrm{H},\mathrm{x}) f(\mathrm{x}|\mathrm{y})\ d\mathrm{x} \right] f(\mathrm{y})\ d\mathrm{y} \tag{25}$$

Now if a loss function that measures the relative importance of various selection errors is specified, the posterior density function is derived from the prior density and the likelihood function $f(y|x)$ of the process is known then the optimal decision can be made for the hypothesis which minimizes the expected loss.

A number of solutions can be obtained to this problem for different loss functions assigned. For instance, a commonly used function for the selection error $(x - \hat{x}^i)$ is its quadratic function, where x is the true parameter vector and $\hat{x}^i$ is the estimated parameter vector for the ith model.

An alternative approach is to use entropy measure of information (Box and Hill, 1967) which has been developed in communication theory (Shannon, 1948).

Consider a complete set of events, E_i, $i = 1, 2,..., m$ whose probabilities are p_1, $p_2,..., p_m$ such that $\sum_{i=1}^{m} p_i = 1$. The expected information (I) of the message on the occurrence of one of these events is defined as

$$I := \sum_{i=1}^{m} p_i \ln \frac{1}{p_i} \tag{26}$$

which is also known as the entropy of the distribution whose probabilities are p_i, $i = 1, 2, ..., m$ (Shannon, 1948). The least possible information occurs when $p_1 = p_2...p_m = 1/m$, which can be derived by maximizing the above equation subject to the constraint $\sum p_i = 1$. In this case, the amount of information is small and the entropy as a measure of disorder is maximum. In other words, all events are equally likely. In situations where the probability of one event p_i is larger than the probability of other events p_j, $j \neq i$ the amount of information is considered to be large and entropy is small. This concept can now be applied to the discrimination of different competing models. Let there be a set of m competing models and the a priori probability of the ith model being true is p_i. If the observations are performed and the a posteriori probability for the ith model is computed then the information gained by this experiment is specified, from (26), as

33

$$\Delta I(H,x) = - \sum_{i=1}^{m} P_{in-1} \ln P_{in-1} + \sum_{i=1}^{m} P_{in} \ln P_{in} \qquad (27)$$

This expression is similar to the previously discussed loss function $L(H,x)$. However in this case the maximum of $\Delta I(H,x)$ is of interest in order to obtain the greatest amount of information out of the experiment. Now, substituting (27) into (22) and considering that there exists a finite number of models, the expected change in entropy (information) $E(\Delta I) =: \Delta J$ before and after the nth observation is obtained

$$\Delta J = \sum_{i=1}^{m} P_{in-1} \int \left\{ \sum_{k=1}^{m} P_{kn} \ln P_{kn} - \sum_{k=1}^{m} P_{kn-1} \ln P_{kn-1} \right\} \cdot p_i(y|x) \, dy_n \qquad (28)$$

where

$$P_{in} = \frac{P_{in-1} \, p_i(y|x)}{\sum_{\ell=1}^{m} P_{\ell n-1} p_\ell(y|x)} \qquad (29)$$

Substituting (29) into (28)

$$\Delta J = \sum_{i=1}^{m} P_{in-1} \int p_i(y|x) \ln \frac{p_i(y|x)}{\sum_{\ell=1}^{m} P_{\ell n-1} p_\ell(y|x)} \, dy_n \qquad (30)$$

Now an observation which maximizes (30) is considered the optimum one. Evaluation of (30) however, is quite complicated but an upper bound for this expression leads to a tractable form. Consider

$$\sum_{j=1}^{m} P_{jn-1} p_i(y|x) \ln \frac{p_i(y|x)}{p_j(y|x)} \geqslant p_i(y|x) \ln \frac{p_i(y|x)}{\sum_{\ell=1}^{m} P_{\ell n} p_\ell(y|x)} \qquad (31)$$

which is a result of Corollary 3.1 of Kullback (1959). Substitution of (31) into (30) gives an upper bound ΔJ_u for ΔJ

$$\Delta J_u = \sum_{i=1}^{m} \sum_{j=1}^{m} P_{in-1} P_{jn-1} \left[\int p_i(y|x) \ln \frac{p_i(y|x)}{p_j(y|x)} \, dy_n + \int p_j(y|x) \ln \frac{p_j(y|x)}{p_i(y|x)} \, dy_n \right]$$

$$(32)$$

34

Let now a group of competing models be given by

$$E\left(y^{(i)}\right) = A^{(i)}x^{(i)} , \quad i = 1, 2, \ldots, m \tag{33}$$

where $y^{(i)}$ is the $m \times 1$ vector of observations for the ith model, $A^{(i)}$ is the nonstochastic $n \times u$ design matrix, and $x^{(i)}$ is the $u \times 1$ unknown parameter vector for the ith model. u is not necessarily the same for all models. If the observations y_n are assumed to be distributed normally with mean $E(y_n)$ and known variance σ^2, the following relationships hold (Box and Hill, 1967)

$$p_i\left(y_n | E(y_n), \sigma\right) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2} \left[y_n - E(y_n)\right]^2\right\} \tag{34}$$

$$p_i\left(E(y_n) | \sigma\right) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left\{-\frac{1}{2\sigma_i^2} \left[E(y_n) - \hat{y}_n^{(i)}\right]^2\right\} \tag{35}$$

where $\hat{y}_n^{(i)}$ is the predicted value of y_n under model i using $n - 1$ observations and its variance is σ_i^2 which is given by

$$\sigma_i^2 = \sigma^2 a_n^{(i)} \left[A^{(i)\prime} A^{(i)}\right]^{-1} a_n^{(i)\prime} \tag{36}$$

where $a_n^{(i)}$ is the row vector of $A^{(i)}$. From the definition of the probability density function of y_n under model i given σ and $n - 1$ observations

$$p_i(y_n | \sigma) = \int p_i\left(y_n | E(y_n), \sigma\right) p_i\left(E(y_n) | \sigma\right) d\left(E(y_n)\right) \tag{37}$$

If (34) and (35) are substituted in (37) and integrated then

$$p_i(y_n | \sigma) = \frac{1}{\sqrt{2\pi(\sigma^2 + \sigma_i^2)}} \exp\left\{-\frac{1}{2(\sigma^2 + \sigma_i^2)} \left(y_n - \hat{y}_n^{(i)}\right)^2\right\} \tag{38}$$

Substitution of (38) into (32) results in the following operational form of discrimination function

$$\Delta J_u = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=i+1}^{m} P_{in-1} P_{jn-1} \left\{ \frac{(\sigma_i^2 - \sigma_j^2)^2}{(\sigma^2 + \sigma_i^2)(\sigma^2 + \sigma_j^2)} + (\hat{y}_n^{(i)} - \hat{y}_n^{(j)})^2 \cdot \right.$$

$$\left. \left(\frac{1}{(\sigma^2 + \sigma_i^2)} + \frac{1}{(\sigma^2 + \sigma_j^2)} \right) \right\} \tag{39}$$

where $\hat{y}_n^{(i)}$ and $\hat{y}_n^{(j)}$ are the predicted observations for model i and j, σ_i^2 and σ_j^2 are the predicted variances obtained from (36) for these predicted observations and σ^2 is the known a priori variance of the observations.

It is now possible to design sequential geodetic surveys to discriminate the descriptive models of deformations using this entropy measure of information. The scenario which is depicted in Figure 2 is as follows.

First an initial network design for the area under consideration is constructed, for instance, using the D-optimal design criteria which is discussed in Section 3.1. Geodetic surveys are then performed at two different epochs covering the whole network. This is followed by the estimation of deformation parameters from the differences of observed quantities. At this point, information provided by the current estimates and prior qualitative information about the models are examined and prior probabilities are assigned to each model. If no preference is inferred from the existing information each model is assigned equal probabilities p = 1/m, where m is the number of models. The next optimal observation (not necessarily the resurvey of the whole network) that gives the maximum expected discrimination among m rival models is sought in (39). Then the new optimal measurement(s) is performed and posteriori probabilities for each model are computed using equation (29). Finally, the current standing of each model is examined. This procedure is repeated each time, using posterior probabilities of previous observations as prior probabilities for the succeeding observations, until one model emerges from the others. The following numerical example illustrates the procedure.

36
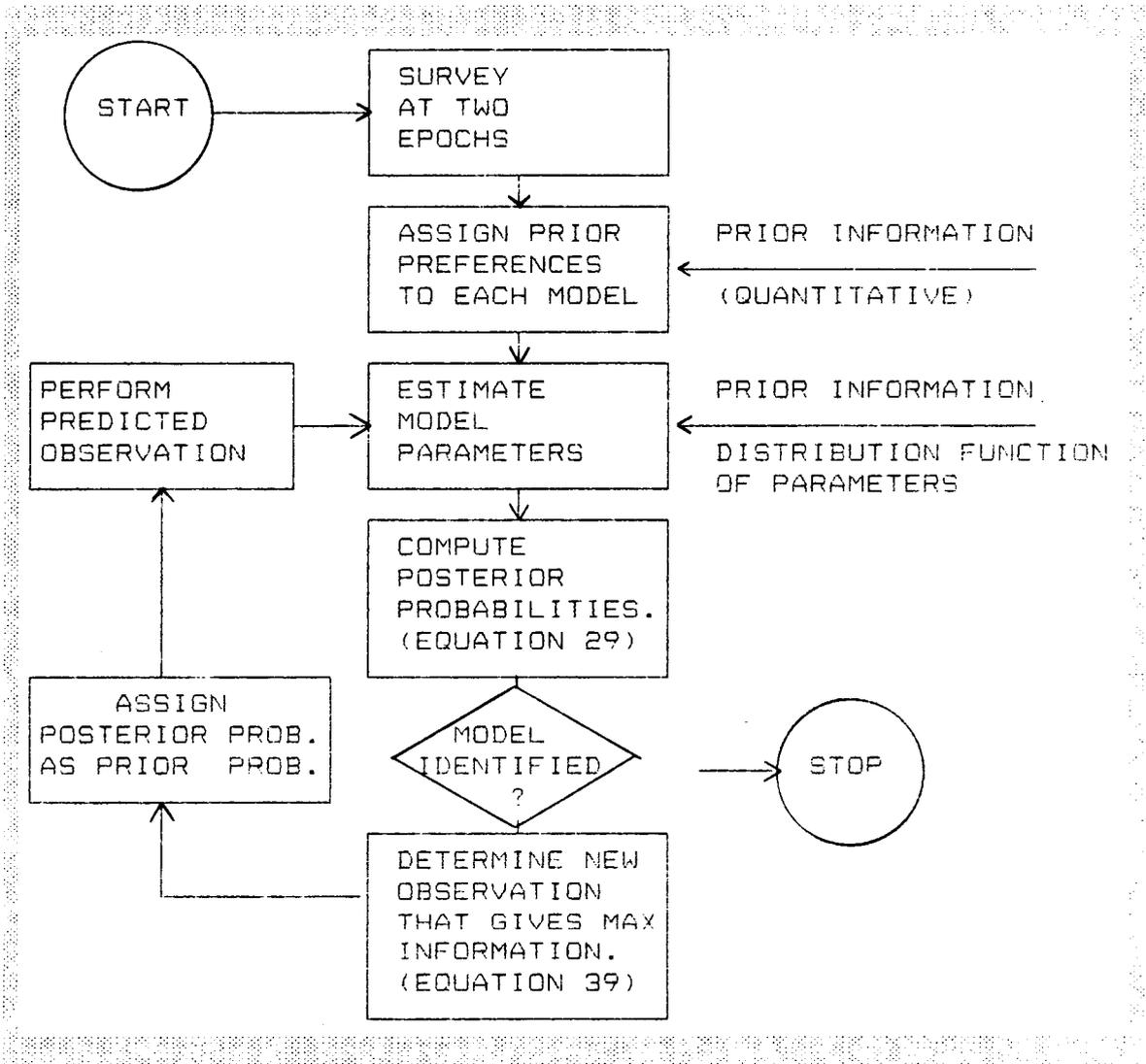
Figure 2. Sequential Model Discrimination

3.3 NUMERICAL EXAMPLE

In Chapter 2, the necessity for sequential experimental designs for the crustal movement analysis was discussed and became part of the proposed algorithm. In the previous section, a method based on the entropy measure of information is presented as a possible candidate for the discrimination of several competing time-dependent models. In order to get a better feeling for the applicability of the method to the crustal deformation analysis, this section describes a numerical example.

Consider the case of homogeneous deformation field. The following descriptive models are postulated as a result of prior experiments to represent possible network deformations,

$$\text{model 1.} \quad \begin{aligned} dx &= (e_x x + e_{xy} y) \, \Delta t \\ dy &= (e_{xy} x + e_y y) \, \Delta t \end{aligned} \tag{40}$$

$$\text{model 2.} \quad dx = e_x x \, \Delta t \tag{41}$$

$$\text{model 3.} \quad \begin{aligned} dx &= e_{xy} y \, \Delta t \\ dy &= e_{xy} x \, \Delta t \end{aligned} \tag{42}$$

where dx and dy are the displacement components of network points for the period Δt, e_x and e_y are the extensional strains in X Y directions. e_{xy} is the shearing strain. If the baselines are observed at different epochs then the baseline length ℓ_{ij} at epoch t is given by the following expression,

$$\ell_{ij}^2 = (x_j - x_i)^2 + (y_j - y_i)^2 \tag{43}$$

Linearizing this expression about the initial epoch t_0 and considering equations (40), (41) and (42) results in the following mathematical models

$$\text{model 1:} \quad \ell_{ijt} - \ell_{ijt_0} = \Delta t \ell_{ij}(\sin^2\alpha_{ij}e_x + \cos^2\alpha_{ij}e_y + \sin 2\alpha_{ij}e_{xy}) \tag{44}$$

$$\text{model 2:} \quad \ell_{ijt} - \ell_{ijt_0} = \Delta t \, \ell_{ij} \sin^2\alpha_{ij} \, e_x \tag{45}$$

$$\text{model 3:} \quad \ell_{ijt} - \ell_{ijt_0} = \Delta t \, \ell_{ij} \sin 2\alpha_{ij} \, e_{xy} \tag{46}$$

38

where ℓ_{ijt} and ℓ_{ijt_0} are the observed baseline lengths at epochs t and t_0 respectively, and α_{ij} is the azimuth of the observed baseline i - j. It is assumed that the scale of baseline observations are the same for different epochs and no stable points or external reference observations are available on the deformable object. Although the differences of the observed baselines can be expressed in terms of the coordinate system invariants of the homogeneous deformation field, the above representation is chosen to investigate and model the contributions of the infinitesimal strain elements in an "adopted" orientation of the coordinate system. Since the above mathematical models (44)-(46) are functions of the baseline azimuths α_{ij}, the orientation of the coordinate system should be kept the same from one epoch to another. Otherwise model parameters are not directly comparable at different epochs.

In this framework, model 1 implies that observed baseline differences are explained by all three components of the infinitesimal strain elements within the adopted orientation of the coordinate system. Model 2 suggests that these differences are only due to extensional strain in x direction, whereas model 3 investigates the differences as a function of the shearing strain. Note that models 2 and 3 are special cases of model 1. For particular orientation of baselines within the adopted coordinate system, models 2 and 3 and models 1 and 2 cannot be differentiated from each other for $\alpha_{ij} = 0°$ and $\alpha_{ij} = 90°$ respectively. However, as we shall see, the discrimination function successfully selects the alternative orientations (namely, the baselines in the N-E and N-W directions) for which all three models are identified.

In this example, model 3 is chosen to be the correct model and the pseudo-observations $(\ell_{ijt} - \ell_{ijt_0})$ are derived using this model. They are also contaminated with a noise from a normal distribution with mean zero and variance 1 mm². Shearing strain, e_{xy}, in the correct model (46) is 0.50 ppm and the time interval between observations is constant and equal to a month. The sequential model discrimination procedure can now be performed using the algorithm depicted in Figure 2.

An initial design is set up for the measurement of deformation parameters (Figure 3a). This is a D-optimal design of model 1, which

39

was derived in Section 3.1, with some additional observations. It is initially postulated that all three models are equally likely. In other words, prior probabilities for each model are 1/3. Part of the network is resurveyed a month later to compute initial estimates for each model's parameters. This makes it possible to predict a new observation for the next month that maximizes the discrimination function ΔJ_u (39) for each model and selects the possible observation which gives the maximum discrimination. In this example three alternatives are possible: baselines which are in the north-east direction, baselines which are in the south-west direction, and baselines which are in the east-west direction. The predicted observation is then performed and posterior probabilities for each model are computed using the new observations and prior probabilities (29).

The rest of the experiment continues following this prediction and observation procedure until computed posterior probabilities indicate that one model is superior to the others (Figures 3c, 3d and 3e). Figure 3 shows that the correct model 3 is identified effectively with 14 baseline observations after 6 months.

As a comparison, null-hypothesis tests (H_0: Model 3 is the same as model 1, H_0: Model 3 is the same as model 2) are performed using the estimated parameters obtained by the least squares method at 5% and 1% significance levels. Neither hypothesis is rejected at 1% significance level until $\Delta t = 5$ months (Figure 3e). In the case of $\alpha = 0.05$, model 2 is rejected at $\Delta t = 5$. Model 2 is rejected at $\Delta t = 6$ months at both levels. However, the results were ambiguous in the sense that model 2 and model 3 were still likely candidates until the last measurement was performed. This problem is clearly eliminated by the proposed method due to the history of accumulated measurements and calculated posterior probabilities of each model.

Meanwhile, the assumptions underlying (39), mainly the normal distribution of observations, will not always be satisfied in practice. Nevertheless, (32) can still be used to derive different criteria for the expected changes in information provided that suitable approximations (distributions) are available for evaluating (32).

40

--- INITIAL DESIGN
    AND OBSERVATIONS

PRIOR PREFERENCES

a)

1/3   1/3   1/3

Model: 1   2   3
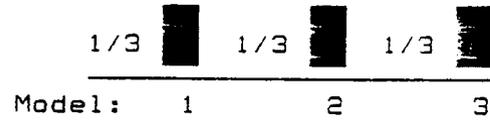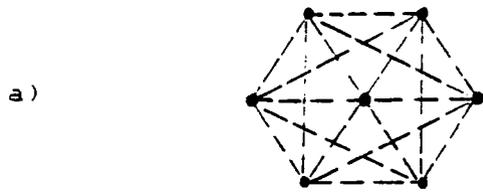
—— REPEATED OBSERVATIONS
▬ PREDICTED OBSERVATION

POSTERIOR PROBABILITIES

b)
Δt=2

0.18   0.50   0.32

1   2   3

c)
Δt=3

0.17   0.37   0.46

1   2   3

d)
Δt=4

0.13   0.21   0.66

1   2   3

e)
Δt=5

0.11   0.04   0.85

1   2   3

f)
Δt=6

0.04   0.00   0.96

1   2   3
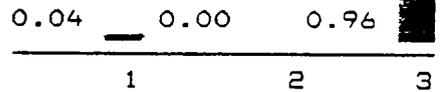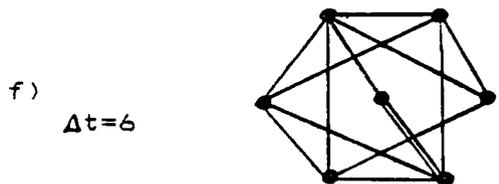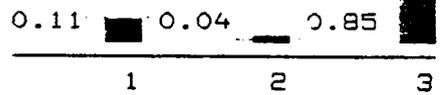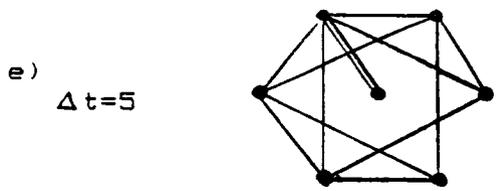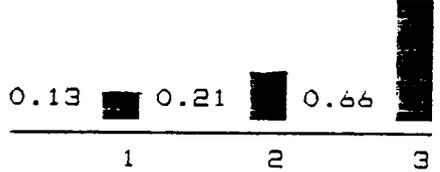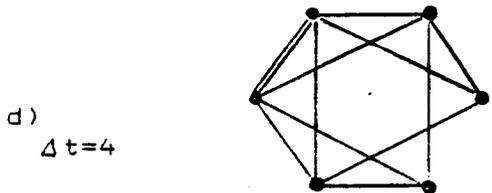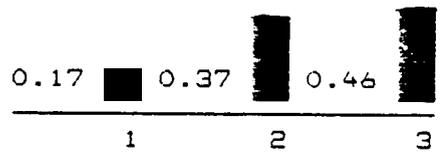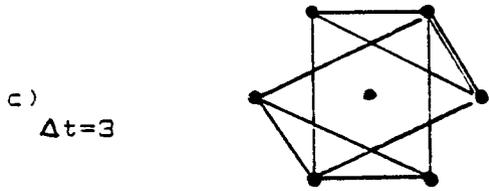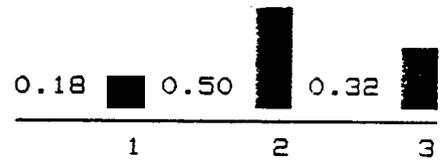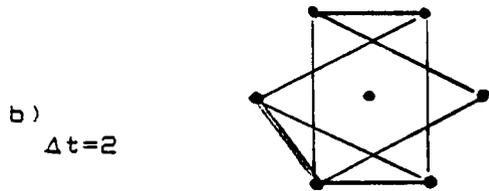
Figure 3.   Results of Sequential Experiment for the
Discrimination of Three Competing Models

41

In addition, H, the set of hypotheses, is assumed to be complete; in other words, the true model is in the set of postulated models in the given example. This is certainly not the case in practice. On the other hand, H may contain a suitable approximation of the true model. In this case the discrimination procedure will require more iterations to identify this approximation. A stopping criterion needs to be established under these circumstances. Also the use of one of the previously outlined model discrimination approaches together with this method will provide some insight to identify a suitable model when more than one model competes after several iterations. If a suitable model cannot be identified, then the discrimination procedure continues, as described in the proposed algorithmic approach, by postulating a new set of models and performing new observations until one model is selected. The methodology presented in this section predicts a single observation for the following epoch. More observations can be predicted using (39). However, in this case, (29) needs to be modified to evaluate the posterior probability of each model.

In any event, the usefulness of the proposed method should be investigated further by applying it to more complicated real world deformation problems.

Chapter 4
ESTIMATION USING PRIOR INFORMATION

4.1 INTRODUCTION

So far, initial design of optimal deformation networks and the experimental design problem for the discrimination of possible crustal movement models has been discussed within the context of the proposed approach. As a result of these steps an operational model which describes the differences of repeatedly measured quantities is identified. Relevant deformation parameters of this model are then estimated using these measurements (sample data). If these estimates are found to be satisfactory, then there is no need to go further. As long as predicted results derived from this model are in agreement with measurements performed at later epochs, it can be used for further investigations. Otherwise the model loses its credibility and the whole procedure starts again to obtain a better model.

Meanwhile uncertainties attributed to these parameters may be large, for instance, due to the precision of repeated observations. In such situations current model estimates can be improved to a certain extent by introducing prior information provided by theoretical models and/or previous investigations about these parameters.

Since this information can be introduced into the estimation process using different types of adjustment techniques, the last part of the proposed algorithm examines some possible alternatives.

Clearly no method of estimation can be perfect, but one method may perform better than the alternatives under a fixed situation. Estimation with prior information is no exception. Our main concern in the proposed algorithm is with the problem of improving the estimation of crustal movement parameters using prior information rather than finding

43

"the best estimator" which will govern the whole analysis under all possible circumstances. It is proposed that different estimators should always be considered for the particular problem under investigation and decision (selection) should be made by assessing the gain for each of the estimators which employs prior information. The results of one estimator which is more efficient than the alternatives are then accepted. Such an approach is obviously inefficient for practical geodetic problems, yet it pays off in crustal movement analysis, where improvements in parameters are very crucial.

In this chapter, first a qualitative criterion to compare different estimators is established. Then, three estimators, the well known mixed model estimation (Best Linear Uniformly Unbiased Estimation BLUUE), Best Linear Estimation (BLE) and Best Linear Unbiased Estimation (BLUE), which all use prior information, are examined under the established criterion.

Since the purpose of using prior information is to improve the estimation of relevant parameters, the above estimators are compared to the Generalized Least Square Estimator (GLSE) which is independent of prior information using the criterion of betterness.

In these comparisons, it is first assumed that observations are affected only by a random noise, that the linear model explains the phenomena and that the prior information is in agreement with the current estimates.

Today controlled instrument calibration procedures, replication and reproduction of observations with different techniques, as well as investigation of different models which fit the data, can guarantee to a certain extent the validity of the first two assumptions. However, possible incompatibilities between prior information and the current data are harder to control and modify. This situation is much more likely to occur in crustal movement analysis simply because the previous data may represent tectonic motions only in the long term average sense or it may be a result of purely theoretical considerations. The existence of these discrepancies can be detected by null-hypothesis testing. However, since the rejection or acceptance of the agreement between

data and extraneous information is also a result of established confidence levels and some additional assumptions, the rejection of the null-hypothesis by no means suggests that such information is totally wrong and should be completely discarded from the analysis.

Therefore, the additional question: "How incorrect can the prior information be so that the resulting estimation with prior information is still advantageous compared to GLSE?" is examined by deriving the necessary conditions for the proposed estimators under biased prior information. A testing procedure is first given for the compatibility of prior and sample information.

In this chapter the following linear model is under consideration

$$y = Ax + u \qquad\qquad (1a)$$

where y is an $n \times 1$ vector of random variables (differences of observed quantities at different epochs). A is an $n \times m$ fixed design matrix of full rank. x is the $m \times 1$ vector of unknown parameters and deterministic in nature. In other words the linear model (1a) is the classical functional relationship model of geodesy. $m < n$ for an overdetermined system. u is the $n \times 1$ vector of disturbances which are random and unobservable with the following assumed properties

$$u \sim (0, \Sigma_{uu}), \quad \Sigma_{uu} > 0 \qquad\qquad (1b)$$

where Σ_{uu} is the $n \times n$ known covariance matrix of disturbances. In addition we will assume that an $m \times 1$ vector of prior information x_0 is available about the parameters x in the following stochastic form

$$x_0 = x + e \qquad\qquad (2a)$$

where e is the $m \times 1$ vector of disturbances with the following distributional properties

$$e \sim (0, \Sigma_{ee}), \quad \Sigma_{ee} > 0 \qquad\qquad (2b)$$

where Σ_{ee} is the known covariance matrix of disturbances. Also we will consider the case where x_0 is the only information available about the parameter vector x in a deterministic form (i.e., when $\Sigma_{ee} \to 0$).

Under this setup, (1)-(2), the general problem is the estimation of the true parameter vector x in the presence of prior information.

4.2 COMPATIBILITY TESTING

Once the model of deformation is identified (discriminatory analysis), the next step is to check whether prior and sample information (i.e., the estimates obtained from the sample observations only) are in agreement with each other. This step can be used as a supporting factor to decide if the prior information should be included in the estimation procedure.

Let x_0 denote the $m \times 1$ vector of prior information on the deformation parameters and let $x_0 = x + e$ where e is the $m \times 1$ error vector, Σ_{ee} its known covariance matrix and x the vector of true parameters. Then the hypothesis H_0: "Prior and sample information are in agreement", can be tested following the similar lines given in (Theil, 1963).

Under this hypothesis, if prior information and an estimate of parameters of GLSE type $\hat{x}_s$ are in agreement, then their difference is expected to be close to zero,

$$\delta := x_0 - \hat{x}_s = e - N^{-1}A'\Sigma_{uu}^{-1}u \tag{3}$$

where $N := A'\Sigma_{uu}^{-1}A$, Σ_{uu} is the $n \times n$ positive definite covariance matrix of observations, u is the corresponding $n \times 1$ error vector, and A is the $n \times u$ design matrix and assumed to be full rank. The matrix of second moments of this difference is

$$E(\delta\delta') = (\Sigma_{ee} + N^{-1}) \tag{4}$$

since $\Sigma_{ee} + N^{-1} > 0$, (positive definite) $E(\delta\delta')$ can be expressed as

$$E(\delta\delta') = Z^{\frac{1}{2}}Z^{\frac{1}{2}} \tag{5}$$

where $Z := (\Sigma_{ee} + N^{-1})$. Then it can be shown that

$$\delta'Z^{-\frac{1}{2}} \sim (0, I). \tag{6}$$

If the disturbance terms on sample observations and prior information are assumed to be distributed normally, i.e.,

$$u \sim N(0, \Sigma_{uu}) \tag{7}$$

$$e \sim N(0, \Sigma_{ee}) \tag{8}$$

then

$$\delta' Z^{-\frac{1}{2}} \sim N(0, I). \tag{9}$$

Therefore, the scalar $\gamma := \delta' Z^{-1} \delta$ can be used as a test statistic since it follows a central χ^2 distribution (Theorem A.10). γ is called the compatibility statistic (Theil, 1963). If α is a predefined error probability of the first kind, then the null-hypothesis will be accepted for $\gamma < \chi^2_{1-\alpha}$.

In the case where x_0 is assumed to be strictly correct ($\Sigma_{ee} \to 0$), the testing procedure is performed using the compatibility statistic which would have the following form

$$\gamma := (x_0 - \hat{x}_s)' N (x_0 - \hat{x}_s) \sim \chi^2_m \tag{10}$$

4.3 COMPARISON CRITERION FOR ESTIMATORS

In comparing two estimators, it is necessary to weigh the advantages of both estimators against their disadvantages. Obviously, such preferences can be expressed in various ways. Each can be best for different purposes and a unique universal criterion unfortunately does not exist for all conceivable situations. Manipulative conveniences of possible alternatives and intuitive arguments about their properties are important in choosing a comparison criterion.

Here the Mean Square Error (MSE) matrix criterion is considered to express qualitative properties of the estimators. This is partly because it gives fairly simple results but mostly because all unbiased estimators turn out to be biased when even one of the assumptions of unbiasedness is violated in practice (Bibby, 1977). This property is especially important in crustal movement analysis when additional

47

information is introduced into the estimation procedure (we consider additional information on the parameters in this study). Since there is no way to guarantee that the additional information is strictly correct, the effect of this information must be exploited for different estimators. Therefore an estimator which is not dominated by the erroneous prior information (i.e., robust against errors in this information) of its alternative can be considered a better estimator.

Also, we expect that this estimator should improve the results in some sense; otherwise there is no reason to introduce additional information into the analysis. The MSE criterion will be used to exploit these types of properties of different estimators.

This choice, however, can be criticized due to the fact that the MSE matrix involves a bias vector which is in practice rarely known and moreover if known can be eliminated. But as it is readily demonstrated in the following sections in comparing different estimators, an unknown bias may still be used as an effective tool in making inferences about the properties of different estimators.

The comparison criterion is defined as follows.

Definition 1: Matrix valued Mean Square Error criterion. An estimator $\hat{x}_1$ of x is said to be better than another estimator $\hat{x}_2$ of x if the difference

$$\text{MSE}(\hat{x}_2) - \text{MSE}(\hat{x}_1) \tag{11}$$

is positive semi-definite (non-negative definite) or equivalently ($\geqslant 0$), where

$$\text{MSE}(\hat{x}_1) = E(x - \hat{x}_1)(x - \hat{x}_1)' \tag{12}$$

$$\text{MSE}(\hat{x}_2) = E(x - \hat{x}_2)(x - \hat{x}_2)' \tag{13}$$

Now let one of the estimators be chosen as the Generalized Least Squares Estimator (GLSE) $\hat{x}_s$ of x which is obviously independent of prior information and is assumed to be unbiased. Let another estimator $\hat{x}$ of x belongs to a class of biased estimators which employs prior information or much more importantly it may be biased as a result of

48

erroneous prior information. Therefore, the corresponding MSE matrices are given by

$$MSE(\hat{x}_s) = D(\hat{x}_s) \tag{14}$$

$$MSE(\hat{x}) = D(\hat{x}) + bb' \tag{15}$$

Here, $D(\hat{x}_s)$ and $D(\hat{x})$ are the $m \times m$ dispersion matrices of $\hat{x}_s$ and $\hat{x}$ respectively and b is the $m \times 1$ vector of bias. Hence, under Definition 1, the estimator $\hat{x}$ which uses prior information is uniformly better than the sample estimator $\hat{x}_s$ of x if

$$MSE(\hat{x}_s) - MSE(\hat{x}) = D(\hat{x}_s) - D(\hat{x}) - bb' \geqslant 0 \tag{16}$$

Since

$$tr\ MSE(\hat{x}_s) = E(x - \hat{x}_s)'(x - \hat{x}_s) \tag{17}$$

$$tr\ MSE(\hat{x}) = E(x - \hat{x})'(x - \hat{x}) \tag{18}$$

the ordinal criterion of betterness given by (16) leads to the following cardinal criterion

$$b'b < tr[D(\hat{x}_s) - D(\hat{x})] \tag{19}$$

The advantages of using the mean square error concept as a comparison criterion should now be clear. (19) indicates that the biased estimator $\hat{x}$ of x is superior to the sample estimator $\hat{x}_s$ of x of GLSE type if its total bias ($b'b$) is less than the total decrease in variances of all estimates. In other words, the use of prior information is justified as long as the total bias introduced remains less than the total improvements of all parameters. This property forms the logical basis in deriving the conditions for the selection of suitable estimators under possibly incompatible prior information.

As we shall see in the following chapters, the mean square error matrix of an estimator usually depends on the true parameter vector x. Consequently, Definition 1 does not exclude the possibility that for some values of x the mean square errors of both estimators will coincide (i.e., $tr\ MSE(\hat{x}_1) = tr\ MSE(\hat{x}_2)$). In these particular cases, there is no basis

49

for preferring one of the estimators over the other, although Definition 1 is still fulfilled. Other properties of the estimators need to be considered.

In the following sections three estimators which employ prior information, namely Best Linear Estimator (BLE), Best Linear Unbiased Estimator (BLUE) and Best Linear Uniformly Unbiased Estimator (BLUUE), are examined using the MSE matrix criterion. They are compared against GLSE under correct and possibly wrong prior information.

4.4 BEST LINEAR ESTIMATION WITH PRIOR INFORMATION

<u>Theorem 1</u>: Best Linear Estimation, BLE, (Tautenburg, 1982; Schaffrin, 1983). Let Gy be a linear estimator of x in the linear model (1). Then the optimum value of G (in the sequel the optimum value of G is denoted with the same letter to simplify the notation) for which

$$E(x - Gy)'(x - Gy) = \text{tr } G\Sigma_{uu}G' + \text{tr}(I - GA)xx'(I - GA)' = \min_G \quad (20)$$

yields to the BLE $\hat{x}_1$ of x

$$\hat{x}_1 := Gy = (1 + x'Nx)^{-1}xx'A'\Sigma_{uu}^{-1}y \quad (21)$$

where

$$N := A'\Sigma_{uu}^{-1}A \quad (22)$$

$$G = (1 + x'Nx)^{-1}xx'A'\Sigma_{uu}^{-1} \quad (23)$$

The dispersion matrix $D(\hat{x}_1)$ of $\hat{x}_1$ is given by

$$D(\hat{x}_1) = (1 + x'Nx)^{-2}x'Nx \cdot xx' \quad (24)$$

Since bias vector b

$$b := x - E(\hat{x}_1) = (I - GA)x \quad (25)$$

is not enforced to be zero in the target function given by (20), the resulting estimator is biased and the bias, by substituting (23) into (25), is

$$b = (I - GA)x \tag{26}$$

$$b = [I - (1 + x'Nx)^{-1}xx'A'\Sigma_{uu}^{-1}A]x \tag{27}$$

$$b = N^{-1}(N^{-1} + xx')^{-1}x \tag{28}$$

and the mean square error matrix $MSE(\hat{x}_1)$ of the estimated parameters, considering (24) and (28), is

$$MSE(\hat{x}_1) = D(\hat{x}_1) + bb' \tag{29a}$$

$$MSE(\hat{x}_1) = (1 + x'Nx)^{-1}xx' \tag{29b}$$

Now let us compare this estimator against the sample estimator $\hat{x}_s$ of GLSE type. The following theorem holds:

<u>Theorem 2</u>: BLE $\hat{x}_1$ of x is better than the sample estimator $\hat{x}_s$ under Definition 4.3.1.

<u>Proof</u>: Consider

$$MSE(\hat{x}_s) = N^{-1} \tag{30}$$

and

$$MSE(\hat{x}_1) = (1 + x'Nx)^{-1}xx' \tag{31}$$

then

$$\Delta := MSE(\hat{x}_s) - MSE(\hat{x}_1) = N^{-1} - (1 + x'Nx)^{-1}xx' \tag{32}$$

$\Delta > 0$ iff

$$(x'Nx)(1 + x'Nx)^{-1} < 1 \tag{33}$$

(Theorem A.1). Since $x'Nx > 0$, (33) is always less than one for $0 < x'Nx < \infty$. This completes the proof.

Although this estimator is better than the sample estimator $\hat{x}_s$, it contains the true parameter vector x and it is applicable only if a valid prior information x_0 about the parameters x such that $x_0 = x$ is

available. This information, however, can not be strictly realized in practice. If it is, then there would be no need for the estimation of parameters. Consequently x_0 to be used in (21) and (29) to replace x in these relationships is in general different from the true parameter vector (i.e., $x_0 \neq x$). In this case it is necessary to compare this estimator versus the sample estimator when the wrong prior information is used in the estimation.

Substituting x_0 for x in (21), <u>the resulting new estimator $\tilde{x}_1$ is</u> given by

$$\tilde{x}_1 = G_0 y = (1 + x_0' N x_0)^{-1} x_0 x_0' A' \Sigma_{uu}^{-1} y \tag{34}$$

where

$$G_0 := (1 + x_0' N x_0)^{-1} x_0 x_0' A' \Sigma_{uu}^{-1} \tag{35}$$

and $x_0 \neq x$. The dispersion matrix $D(\tilde{x}_1)$ of $\tilde{x}_1$ is now given by

$$D(\tilde{x}_1) = G_0 \Sigma_{uu}^{-1} G_0' = (1 + x_0' N x_0)^{-2} x_0' N x_0 \cdot x_0 x_0' \tag{36}$$

and the bias b_0 is

$$b_0 := x - E(\tilde{x}_1) = (I - G_0 A)x$$
$$b_0 = N^{-1}(N^{-1} + x_0 x_0')^{-1} x. \tag{37}$$

The mean square error matrix of $\tilde{x}_1$ now has the following form

$$MSE(\tilde{x}_1) = D(\tilde{x}_1) + b_0 b_0' = (1 + x_0' N x_0)^{-2} x_0' N x_0 \cdot x_0 x_0' + N^{-1}(N^{-1} + x_0 x_0')^{-1}$$
$$\cdot xx'(N^{-1} + x_0 x_0')^{-1} N^{-1} \tag{38}$$

Comparing (38) with the MSE matrix of the sample estimator $\hat{x}_s$ which is obviously not affected by any type of prior information results in the following relationship

$$MSE(\hat{x}_s) - MSE(\tilde{x}_1) = N^{-1} - D(\tilde{x}_1) - b_0 b_0' \tag{39}$$
$$= N^{-1} - (1 + x_0' N x_0)^{-2} x_0' N x_0 \cdot x_0 x_0' - N^{-1}(N^{-1} + x_0 x_0')^{-1}$$
$$\cdot xx'(N^{-1} + x_0 x_0')^{-1} N^{-1} \tag{40}$$

It can be shown by using Theorem A.1 that the difference of the first two terms in the above equation is always positive definite. Now let

$$B := N^{-1} - D(\tilde{x}_1) = N^{-1} - (1 + x_o'Nx_o)^{-2}x_o'Nx_o \cdot x_ox_o' \tag{41}$$

then using Theorem A.1 again, (39) is p.s.d. iff

$$x'N^{-1}(N^{-1} + x_ox_o')^{-1}B^{-1}(N^{-1} + x_ox_o')^{-1}N^{-1}x \leqslant 1 \tag{42}$$

This expression reduces after some simple but lengthy manipulations to

$$x'(N^{-1} + 2x_ox_o')^{-1}x \leqslant 1 \tag{43}$$

Hence the following theorem is proven.

<u>Theorem 3</u>: The necessary and sufficient condition for the biased estimator $\tilde{x}_1$ under wrong prior information to be better than the sample estimator $\hat{x}_s$ for all $x_o \neq x$ with respect to Definition 4.3.1 is

$$x'(N^{-1} + 2x_ox_o')^{-1}x \leqslant 1 \tag{44}$$

The implications of this theorem will be examined at the end of this chapter together with the other theorems which will be derived in the following sections.

Let us now examine the behavior of bias introduced by the use of wrong prior information when the above condition holds. The quadratic bias, from (37), is given by

$$b_o'b_o = x'(N^{-1} + x_ox_o')^{-1}N^{-1}N^{-1}(N^{-1} + x_ox_o')^{-1}x \tag{45}$$

which can be expressed, after a simple transformation, as

$$b_o'b_o = z'(I + z_oz_o')^{-1}N^{-1}(I + z_oz_o')^{-1}z \tag{46}$$

where

$$z := N^{\frac{1}{2}}x \tag{46a}$$
$$z_o := N^{\frac{1}{2}}x_o \tag{46b}$$

Hence, the condition for betterness under erroneous prior information (Theorem 3) takes the following form

$$z'(I + 2z_0z_0')^{-1}z \leqslant 1 \tag{47}$$

The following corollary holds.

Corollary 1: The quadratic bias due to the incompatible prior information under the condition of betterness (44) is bounded. The supremum of the total (quadratic) bias as a result of the wrong prior information is equal to or less than the largest eigenvalue λ_m of N^{-1}, i.e.,

$$\sup \quad z'(I + z_0z_0')^{-1}N^{-1}(I + z_0z_0')^{-1}z \leqslant \lambda_m \tag{48}$$

$$\text{s.t.:} \quad z'(I + 2z_0z_0')^{-1}z \leqslant 1 \tag{49}$$

Proof: Let $N^{-1} = \Gamma\Lambda\Gamma'$ be a spectral decomposition of the symmetric matrix N^{-1} and also let

$$d: = \Gamma(I + z_0z_0')^{-1}z \tag{50}$$

then

$$d'd = z'(I + z_0z_0')^{-1}(I + z_0z_0')^{-1}z \tag{51}$$

so that

$$\sup \ b_0'b_0 = \sup \ d'\Lambda d = \sup_{i=1}^{m} \lambda_i d_i^2 \tag{52}$$

where λ_i are the eigenvalues of N^{-1}. If these eigenvalues are written in descending order then the above equation (52) satisfies

$$\sup \sum_{i=1}^{m} \lambda_i d_i^2 \leqslant \lambda_m \sup \sum_{i=1}^{m} d_i^2 \tag{53}$$

Using Theorem A.7, it can be shown that

$$z'(I + 2z_0z_0')^{-1}z - d'd \geqslant 0 \tag{54}$$

54

therefore

$$\lambda_m \sup \sum_{i=1}^{m} d_i^2 \leqslant \lambda_m \sup z'(I + 2z_0 z_0')^{-1} z \leqslant \lambda_m \tag{55}$$

since $z'(I + 2z_0 z_0')^{-1} z \leqslant 1$. This completes the proof.

It should be noted here that the boundedness in the quadratic bias also implies the boundedness of its elements (Voievodine, 1976).

4.5 BEST LINEAR UNBIASED ESTIMATION WITH PRIOR INFORMATION

In this section another estimator, namely Best Linear Unbiased Estimator (BLUE), which is recently derived by Schaffrin (1983), is examined. The estimator and the relevant statistics are stated. Since the purpose of using such an estimator, in the scope of the proposed algorithm, is to improve the estimates, it is compared against GLSE. This estimator, similar to BLE, requires additional information x_0 about the parameters to be estimated. Comparisons are also made for the case when the prior information is wrong. The main difference between BLE and BLUE is that the unbiasedness condition is weakly introduced into the target function given by (20) which appears to be a novel in statistical literature.

Theorem 1: Best Linear Unbiased Estimation, BLUE, (Schaffrin, 1983). Let Gy be a linear estimator of x in the linear model (1). Then the optimum value of G for which

$$\text{tr } D(Gy) = \text{tr } G \Sigma_{uu} G' = \min_{G} \tag{56}$$

subject to

$$b := x - E(Gy) = (I - GA)x = 0 \tag{57}$$

in the sense of,

$$\text{tr}[G\Sigma_{uu}G' - 2(I - GA)x\lambda'] = \min_{G} \tag{58}$$

55

where λ is a Lagrange multiplier vector and $D(Gy)$ is the dispersion matrix of Gy, yields to the BLUE $\hat{x}_2$ of x

$$\hat{x}_2 := Gy = (x'Nx)^{-1}xx'A'\Sigma_{uu}^{-1}y \tag{59}$$

where

$$G = (x'Nx)^{-1}xx'A'\Sigma_{uu}^{-1} \tag{60}$$

$$N := A'\Sigma_{uu}^{-1}A \tag{61}$$

The dispersion matrix $D(\hat{x}_2)$ of $\hat{x}_2$ is given by

$$D(\hat{x}_2) = (x'Nx)^{-1}xx' \tag{62}$$

Since the bias b

$$b := x - E(\hat{x}_2) \tag{63}$$

is enforced to be zero in the target function (56), the resulting estimator turns out to be unbiased, that is

$$E(\hat{x}_2) = E\left[(x'Nx)^{-1}xx'A'\Sigma_{uu}^{-1}y\right] = (x'Nx)^{-1}xx'A'\Sigma_{uu}^{-1}Ax = x \tag{64}$$

Therefore, the mean square error matrix of the estimate is

$$MSE(\hat{x}_2) = D(\hat{x}_2) \tag{65}$$

Since the unbiasedness condition is introduced through (57) which has a solution only for the scalar multiples of x, the resulting estimator is weakly unbiased (see (Schaffrin, 1983) for the general concept of unbiasedness). If the condition of unbiasedness is introduced through the homogeneous equation $(I - GA) = 0$ which obviously holds for any x, then the resulting estimator is the well-known best linear uniformly unbiased estimator (or equivalently GLSE) which is independent of prior information.

Note that this estimator reduces to GLSE for the univariate case (i.e., $m = 1$).

<u>Proposition 2</u>: BLUE $\hat{x}_2$ of x is better than the sample estimator $\hat{x}_s$ of x with respect to Definition 4.3.1.

<u>Proof</u>: Consider the differences of the dispersion matrices of both estimators

$$\Delta := D(\hat{x}_s) - D(\hat{x}_2) \tag{66}$$

then from (62)

$$\Delta = N^{-1} - (x'Nx)^{-1}xx' \tag{67}$$

which can also be written as

$$\Delta = N^{-\frac{1}{2}}[I - N^{\frac{1}{2}}(x'Nx)^{-1}xx'N^{\frac{1}{2}}]N^{-\frac{1}{2}} \tag{68}$$

therefore $\Delta \geqslant 0$ iff

$$I - (x'Nx)^{-1}N^{\frac{1}{2}}xx'N^{\frac{1}{2}} \geqslant 0 \tag{69}$$

which holds as a result of Theorem A.1. This completes the proof.

Similar to BLE, this estimator is a function of the true parameter vector x and it is applicable if valid prior information $x_0 \propto x$ (x_0 is proportional to x, i.e., $x_0 = c \cdot x$ where c is a real number) is available. Again this condition cannot be achieved in practice. Therefore the resulting estimates are no longer unbiased since $x_0 \neq x$ (x_0 is not proportional to x) in general. <u>This biased estimator $\tilde{x}_2$ is now given by</u>

$$\tilde{x}_2 = G_0 y = (x_0'Nx_0)^{-1}x_0x_0'A'\Sigma_{uu}^{-1}y \tag{70}$$

where

$$G_0 = (x_0'Nx_0)^{-1}x_0x_0'A'\Sigma_{uu}^{-1} \tag{71}$$

with $x_0 \neq x$. The dispersion matrix $D(\tilde{x}_2)$ of $\tilde{x}_2$ is, from (70)

$$D(\tilde{x}_2) = (x_0'Nx_0)^{-1}x_0x_0' \tag{72}$$

and the bias b_0 caused by the wrong prior information is

$$b_0 := x - E(\tilde{x}_2) = (I - G_0A)x = \left(I - \frac{x_0 x_0' N}{x_0' N x_0}\right) x \tag{73}$$

It is clear that the dispersion matrix of the estimate does not fully measure the effect of the bias caused by the introduction of erroneous prior information. We therefore consider Definition 4.3.1. Considering (72) and (73)

$$\Delta := MSE(\hat{x}_s) - MSE(\tilde{x}_2) = D(\hat{x}_s) - D(\tilde{x}_2) - b_0 b_0'$$

$$\Delta = B - BNxx'NB \tag{74}$$

where

$$B := N^{-1} - (x_0' N x_0)^{-1} x_0 x_0' \tag{75}$$

Since $B \geqslant 0$ (Theorem 2), it can be represented as $B = B^{\frac{1}{2}} B^{\frac{1}{2}}$ (Theorem A.2), then

$$\Delta = B^{\frac{1}{2}}(I - B^{\frac{1}{2}}Nxx'NB^{\frac{1}{2}})B^{\frac{1}{2}} \tag{76}$$

and this expression is p.s.d. iff

$$I - B^{\frac{1}{2}}Nxx'NB^{\frac{1}{2}} \geqslant 0 \tag{77}$$

Using Theorem A.1, (77) is p.s.d. iff

$$x'NBNx \leqslant 1 \tag{77a}$$

which can also be expressed, substituting (75), as

$$x'Nx - \frac{(x'Nx_0)^2}{x_0'Nx_0} \leqslant 1 \tag{78}$$

To explore the behavior of the above inequality under varying wrong prior information, consider the extrema of the second quadratic form

$$\sup_{x_0} \frac{(x'Nx_0)^2}{x_0'Nx_0} = x'Nx \tag{79}$$

as a result of Theorem A.3. This is the case when the prior information is correct. On the other hand

$$\inf_{x_0} \frac{(x'Nx_0)^2}{x_0'Nx_0} = \inf_{z_0} \frac{z_0'Cz_0}{z_0'z_0} = \lambda_m \tag{80}$$

as a result of Theorem A.4, where

$$C := N^{\frac{1}{2}}xx'N^{\frac{1}{2}} \tag{81}$$

$$z_0 = N^{\frac{1}{2}}x_0 \tag{82}$$

and λ_m is the minimum eigenvalue of C. Since rank(C) = 1, $\lambda_m = 0$. Therefore

$$\inf_{x_0} \frac{(x'Nx_0)^2}{x_0'Nx_0} = 0 \tag{83}$$

which, together with (79), leads to the following theorem.

<u>Theorem 3</u>: A sufficient condition for the biased estimator $\tilde{x}_2$ to be better than the sample estimator $\hat{x}_s$ for all $x_0 \neq x$ <u>with respect to Definition 4.3.1</u> is

$$x'Nx \leq 1 \tag{84}$$

<u>Corollary 1</u>: If $u \sim N(0, \sigma_u^2 I)$, then a sufficient condition for the biased estimator $\tilde{x}_2$ to be better than $\hat{x}_s$ for all $x_0 \neq x$ is

$$\frac{1}{\sigma_u^2} x'A'Ax \leq 1 \tag{85}$$

which holds basically when the signal to noise ratio

$$\sum_{i=1}^{m} \frac{x_i^2}{\sigma_u^2} \tag{86}$$

59

is sufficiently small.

Again leaving further implications of the above results to the end of this chapter, let us examine the amount of bias committed by the use of incompatible additional information. From (73) and (82) the quadratic bias is given by

$$b_0'b_0 = z'\left[I - \frac{z_0 z_0'}{z_0' z_0}\right]N^{-1}\left[I - \frac{z_0 z_0'}{z_0' z_0}\right]z \tag{87}$$

and the condition of betterness (84), after a simple transformation, is

$$z'z \leqslant 1 \tag{88}$$

where $z := N^{1/2}x$. Now observe that

$$z'\left[I - \frac{z_0 z_0'}{z_0' z_0}\right]\left[I - \frac{z_0 z_0'}{z_0' z_0}\right]z = z'\left[I - \frac{z_0 z_0'}{z_0' z_0}\right]z \tag{89}$$

and

$$\sup_{z_0} z'\left[I - \frac{z_0 z_0'}{z_0' z_0}\right]z = z'z \tag{90}$$

Considering the spectral decomposition of N^{-1}

$$N^{-1} = C\Lambda C' \tag{91}$$

where C is an orthogonal matrix and Λ is the m × m diagonal matrix of λ_i eigenvalues of N^{-1}, equation (87) can be written as

$$b_0'b_0 = d'C\Lambda C'd \tag{92}$$

where

$$d' := z'\left[I - \frac{z_0 z_0'}{z_0' z_0}\right] \tag{93}$$

Now, from (92)

$$b_0'b_0 = \sum_{i=1}^{m} \lambda_i d_i^2 . \tag{94}$$

then

$$\sup_{z_0} b_o' b_o = \sup \sum_{i=1}^{m} \lambda_i d_i^2 \leqslant \lambda_m \sup \sum_{i=1}^{m} d_i^2 \tag{95}$$

where λ_m is the largest eigenvalue of N^{-1}. Substituting (89) and (90) into the last term in (95) and considering (88)

$$\sup_{\substack{z'z \leqslant 1}} b_o' b_o \leqslant \lambda_m \tag{96}$$

is obtained. Therefore, the following corollary is proven.

Corollary 2: The total bias due to the wrong prior information is bounded under the condition of improvement given by (84). The supremum of it is equal to or less than the largest eigenvalue of N^{-1}, i.e.,

$$\sup b_o' b_o \leqslant \lambda_m \tag{97}$$
$$s.t.: x'Nx \leqslant 1$$

4.6 BEST LINEAR UNIFORMLY UNBIASED ESTIMATION WITH PRIOR INFORMATION

In the previous two sections, the two estimators examined require prior information on the parameters to be known in an exact sense. There may be situations, however, that an independent "stochastic" auxiliary information vector x_0 on x is available, i.e.,

$$x_0 = x + e$$
$$e \sim (0, \Sigma_{ee}) , \quad \Sigma_{ee} > 0 \tag{98}$$

where e is the $m \times 1$ vector of disturbances on x_0 and Σ_{ee} is the $m \times m$ covariance matrix. The availability of this information leads to the following mixed model

$$\bar{y} = \bar{A}x + \bar{u} \tag{99}$$

where

$$\bar{y} := \begin{bmatrix} y \\ x_0 \end{bmatrix} \quad , \quad \bar{A} := \begin{bmatrix} A \\ I \end{bmatrix} \tag{100}$$

$$\bar{u} := \begin{bmatrix} u \\ e \end{bmatrix} \sim \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} , \begin{bmatrix} \Sigma_{uu} & 0 \\ 0 & \Sigma_{ee} \end{bmatrix} =: \Sigma_{\bar{u}\bar{u}} \right\} \tag{101}$$

Here y, A, x, u and Σ_{uu} have the same meaning as defined in the linear model given by (1). The problem of estimation in this mixed model is well known. In this section first the Best Linear Uniformly Unbiased Estimation (BLUUE) technique is summarized. It is then compared against the GLSE. Results are due to Terasvirta (1979). However, the amount of bias committed by the use of extraneous information given by Corollary 2 seems to be novel.

<u>Theorem 1</u>: Best Linear Uniformly Unbiased Estimation, (Theil and Goldberger, 1961). Consider the linear model given by (98) – (101). Let $\bar{G}\bar{y}$ be a linear estimator of x in this model, then the optimum value of $\bar{G}$ for which

$$\text{tr } D(\bar{G}\bar{y}) = \text{tr } \bar{G}\Sigma_{\bar{u}\bar{u}}\bar{G}' = \min_{\bar{G}} \tag{102}$$

$$\text{s.t.:} \quad (I - \bar{G}\bar{A}) = 0 \tag{103}$$

for all x, yields to the BLUUE $\hat{x}_3$ of x

$$\hat{x}_3 = (I + \Sigma_{ee}N)^{-1}(\Sigma_{ee}A'\Sigma_{uu}^{-1}y + x_0) \tag{104}$$

whose dispersion matrix is given by

$$D(\hat{x}_3) = (\Sigma_{ee}^{-1} + N)^{-1} \tag{105}$$

where $N := A'\Sigma_{uu}^{-1}A$.

Since the effect of additional information is expected to improve the estimates, the following theorem is of interest.

Theorem 2: The BLUUE $\hat{x}_3$ of x is a better estimate $\hat{x}_s$ of x in the sense of Definition 4.3.1.

Proof: Since both estimators are unbiased, Definition 4.3.1 turns out to be a variance comparison. From (105)

$$D(\hat{x}_s) - D(\hat{x}_3) = N^{-1} - (\Sigma_{ee}^{-1} + N)^{-1} \tag{106}$$

This expression is p.s.d. as a result of Theorem A.7. This completes the proof.

The unbiasedness assumption, however, can hardly be realized in practice. The compatibility of auxiliary information with the information implied by the observations is especially difficult to achieve. Since the current observations are performed under controlled circumstances, it is expected that incompatibility is more likely to be due to the additional information.

Assume that the true but unknown form of the additional information on the parameters is in fact

$$x_0 = x + s + e \tag{107}$$

where $s \neq 0$ is an m×1 vector of possible deviations (systematic errors) and deterministic in nature. This in turn implies that, from (104)

$$\tilde{x}_3 = (I + \Sigma_{ee}N)^{-1}(\Sigma_{ee}A'\Sigma_{uu}^{-1}y + x_0) \tag{108}$$

where x_0 is now given by (107). Since

$$E(x_0) \neq x \tag{109}$$

this estimator is no longer unbiased, and it will be denoted by $\tilde{x}_3$ to emphasize the bias. It is easy to show, using (107) and (108), that

$$E(\tilde{x}_3) = x + (I + \Sigma_{ee}N)^{-1}s \tag{110}$$

Hence the last term is the bias caused by the wrong prior information

$$b = (I + \Sigma_{ee}N)^{-1}s \tag{111}$$

63

Now the next question is, "under what conditions can wrong prior information still be used to improve the estimates?". Following Definition 4.3.1, the biased estimator $\tilde{x}_3$ dominates the sample estimate $\hat{x}_s$ iff

$$MSE(\hat{x}_s) - MSE(\tilde{x}_3) \geqslant 0 \tag{112}$$

$$tr[MSE(\hat{x}_s) - MSE(\tilde{x}_3)] \geqslant 0 \tag{113}$$

$$tr[D(\hat{x}_s) - D(\tilde{x}_3)] - s'B'Bs \geqslant 0 \tag{114}$$

where $B := (I + \Sigma_{ee}N)^{-1}$. Now using Theorems A.1 and A.9, (113) and (112) hold iff

$$s'(\Sigma_{ee} + N^{-1})^{-1}s \leqslant 1 \tag{115}$$

This form leads to the following corollary.

Corollary 1: (Terasvirta, 1979). A necessary and sufficient condition for the biased estimator $\tilde{x}_3$ to be superior than the sample estimator $\hat{x}_s$ in the sense described by Definition 4.3.1 is

$$s'(\Sigma_{ee} + N^{-1})^{-1}s \leqslant 1 \tag{116}$$

Let us now examine the amount of bias introduced by the use of extraneous information under the above improvement condition.

Corollary 2: The total bias due to the wrong prior information under the condition of improvement (116) is bounded. The supremum of it is equal to or less than the largest eigenvalue λ_m of $(I + N\Sigma_{ee})^{-1}N^{-1}$, i.e.,

$$\sup_{x_0} b'b \leqslant \lambda_m \tag{117}$$
$$\text{s.t.:} \quad s'(\Sigma_{ee} + N^{-1})^{-1}s \leqslant 1$$

Proof: After a simple transformation the total bias, from (111), and the condition of improvement, from (116), can be written as

$$b'b = z_0'(\Sigma_{ee} + N^{-1})^{-\frac{1}{2}}N^{-1}N^{-1}(\Sigma_{ee} + N^{-1})^{-\frac{1}{2}}z_0 \tag{118}$$

64

$$z_0'z_0 \leqslant 1 \tag{119}$$

where $z_0' := s'(\Sigma_{ee} + N^{-1})^{-\frac{1}{2}}$ then,

$$\sup_{z_0} b'b = \lambda_m$$
$$\text{s.t.:} \quad z_0'z_0 \leqslant 1 \tag{120}$$

as a result of Theorem A.4, where λ_m is the largest eigenvalue of

$$(\Sigma_{ee} + N^{-1})^{-\frac{1}{2}}N^{-1}N^{-1}(\Sigma_{ee} + N^{-1})^{-\frac{1}{2}} \tag{121}$$

Since

$$\text{tr}\left[(\Sigma_{ee}+N^{-1})^{-\frac{1}{2}}N^{-1}N^{-1}(\Sigma_{ee}+N^{-1})^{-\frac{1}{2}}\right] = \text{tr}\left[(I+N\Sigma_{ee})^{-1}N^{-1}\right] = \sum_{i=1}^{m}\lambda_i \tag{122}$$

the largest eigenvalue of (122) is equal to the largest eigen-value of $(I + N\Sigma_{ee})^{-1}N^{-1}$. This completes the proof.

4.7 FURTHER DISCUSSIONS

In the previous sections, three different estimators which employ prior information were examined. Since the purpose of using prior information about the parameters is to improve the sample estimates (i.e., the estimates which are obtained without priors) according to the proposed algorithmic approach, they were compared against the GLSE technique in order to determine the effect of this information. Comparisons were made using the MSE matrix criterion. It was shown that introduction of prior information improves the results according to the MSE matrix criterion (except BLUE which reduces to GLSE in the case of univariate application) for the case when prior information is strictly correct. Comparisons were also made when prior information is not compatible which is the case in practical applications. Improvement conditions over GLSE were derived for this case. Results are summarized in Table 1.

In this section these results are interpreted within the scope of the proposed algorithmic approach. First these estimators are compared

TABLE 1: ESTIMATORS WITH PRIOR INFORMATION

ESTIMATOR	BLE	BLUE	BLUUE
ESTIMATOR	$\hat{x}_1 = \left[(1+x_o'Nx_o)^{-1}x_o'A'\Sigma_{uu}^{-1}y\right]\cdot x_o$	$\hat{x}_2 = \left[(x_o'Nx_o)^{-1}x_o'A'\Sigma_{uu}^{-1}y\right]\cdot x_o$	$\hat{x}_3 = (I+\Sigma_{ee}N)^{-1}(\Sigma_{ee}A'\Sigma_{uu}^{-1}y+x_o)$
IMPROVEMENT CONDITIONS WHEN x_o IS NOT COMPATIBLE	$x'(N^{-1}+2x_ox_o')^{-1}x \leqslant 1$	$x'Nx \leqslant 1$	$s'(\Sigma_{ee}+N^{-1})^{-1}s \leqslant 1$
TOTAL BIAS INTRODUCED	$b_o'b_o \leqslant \lambda_m \in N^{-1}$	same as BLE	$b_o'b_o \leqslant \lambda_m \in (I+N\Sigma_{ee})^{-1}N^{-1}$
ASSUMPTIONS	$u \sim (0,\sigma_u^2I)$	same as BLE	$u \sim (0,\sigma_u^2I),\; e \sim (0,\sigma_e^2I)$
CONTROL VARIABLES FOR IMPROVEMENTS	$\dfrac{x'A'Ax}{\sigma_u^2} < 1$	same as BLE	$\displaystyle\sum_{i=1}^{m} \frac{t_i^2/\sigma_u^2}{k^{-1}+\lambda_i} \leqslant 1$ $(A'A)^{-1} = C\Lambda C'$, $\quad t = C's$ $k = \sigma_u^2/\sigma_e^2$, $\quad s = x - E(x_o)$

N: $= (A'\Sigma_{uu}^{-1}A)^{-1}$

with respect to each other under correct prior information; then it was shown that direct comparisons are not possible when prior information is not compatible. In this case their corresponding improvement conditions can be used to select a suitable estimator.

Now consider BLE, from (21)

$$\hat{x}_1 = (1 + x'Nx)^{-1} xx'A'\Sigma_{uu}^{-1} y \tag{123}$$

This estimator is better than GLSE (Theorem 4.4.2) but not practical as it is, since it is a function of the true parameter vector x. To overcome this difficulty several approaches are given in statistical literature. For instance, Farebrother (1976) replaces x with $\hat{x}_s$ of GLSE type. Vinod (1976) iterates the Farebrother estimator and gives an analytic solution to this iterative form. Another estimator results by postulating the linear model (1) as $y = Ax + ku$ where k is an arbitrary number and modifying (123) accordingly. Ullah and Ullah (1978) introduce in (123) two constants k_1, k_2 where $k_1 > 0$, k_2 are arbitrary stochastic or nonstochastic scalars (double k-class estimators). They also show that for particular values of k_1 and k_2 this new estimator reduces to a James and Stein Estimator (1961).

Common in all the above approaches is the use of sample information. However if valid prior information x_0 which is independent of the current observations is available about the true values of the parameters, then (123) can readily be exploited by substituting x_0 for x. In this case the usefulness of the resulting estimator depends on how close x_0 is to x. In the meantime since the purpose of introducing prior information, according to the proposed approach, is to improve the estimates, the introduction of prior information is not only helpful in reducing the MSE of estimates in an algebraic sense but also meaningful in the sense that this information may contain a component of reality which is not captured by the sample observations.

If this information x_0 is available, then (123) can be written as (we denoted this estimator as $\hat{x}_1$ when $x_0 = x$, and $\tilde{x}_1$ when $x_0 \neq x$ in the previous sections)

$$\hat{x}_1 = (1 + x_0' N x_0)^{-1} x_0 x_0' \ A' \Sigma_{uu}^{-1} \ y \tag{124}$$

For the univariate case (124) reduces to

$$\hat{x}_1 = \alpha \cdot \hat{x}_s \tag{125}$$

where

$$\alpha := \left[1 + \frac{\sigma_u^2}{x_0^2 n} \right]^{-1}, \quad 0 < \alpha < 1 \text{ for all } x_0 \tag{126}$$

which indicates that this estimator provides a shrinkage of the sample estimator $\hat{x}_s$ of GLSE type, and since

$$E(\hat{x}_1) = \alpha \cdot x \tag{127}$$

it also underestimates the true parameter vector on average, and it is therefore a biased estimator. Yet it is better than the GLSE (Theorem 4.4.2) provided that x_0 is strictly correct (i.e., $x_0 = x$). Obviously this information is not known strictly in practice. In this case (when $x_0 \neq x$), uniform improvements over GLSE by using x_0 cannot be guaranteed. Nevertheless it is still useful if the general improvement condition which was derived in section 4.4 holds. The following considerations lead to simpler interpretations. From (44)

$$x' \ (N^{-1} + 2x_0 x_0')^{-1} x \leqslant 1 \tag{128}$$

Since

$$N - (N^{-1} + 2x_0 x_0')^{-1} > 0 \qquad \text{for } x_0 \neq 0 \tag{129}$$

an upper bound for (128) with respect to x_0 is

$$x' N x < 1 \tag{130}$$

For the sake of simplicity assume that disturbances are

$$u \sim (0, \ \sigma_u^2 I) \tag{131}$$

Substituting (131) into (130) and using the spectral decomposition of $(A'A)^{-1}$, (130) can be replaced by

$$\sum_{i-1}^{m} \frac{t_i^2/\sigma_u^2}{\lambda_i} < 1 \tag{132}$$

where λ_i are the eigenvalues of $(A'A)^{-1}$, t_i are the elements of $t = C'x$. This expression indicates that the prior information will improve the results when

(i)   The signal-to-noise ratio $(x'x/\sigma_u^2)$ is sufficiently small.

(ii)   The component of signal x is not very pronounced when λ_i is small.

Consider now BLUE given by (59) and which can also be expressed as

$$\hat{x}_2 = \frac{xx'N}{x'Nx} \hat{x}_s \tag{133}$$

Since the elements of $xx'N/x'Nx$ are bounded, BLUE displays similar properties as BLE.   Again prior information x_0 about x needs to be known in order for this estimator to be applicable.   In addition, it is unbiased if x_0 is proportional to x $(x_0 \propto x)$.   However, it reduces to GLSE for the univariate case and thereby offers no improvements.   If BLUE is compared to BLE when $x_0 = x$ under Definition 4.3.1

$$\text{MSE}(\hat{x}_2) - \text{MSE}(\hat{x}_1) = D(\hat{x}_2) - \text{MSE}(\hat{x}_1) = (x_0'Nx_0)^{-1}(1 + x_0'Nx_0)^{-1}x_0x_0' \tag{134}$$

is obtained.   This expression is p.s.d. due to its quadratic nature. BLUE is therefore not as efficient as BLE over GLSE under strictly correct prior information.   On the other hand, if this information is not strictly correct then it is rather difficult to establish the superiority of one estimator over the other by examining their corresponding MSE matrices.   From (38), (39) and (68), (69).

$$\text{MSE}(\tilde{x}_2) - \text{MSE}(\tilde{x}_1) = D(\tilde{x}_2) - D(\tilde{x}_1) + b_2b_2' - b_1b_1' \tag{135}$$

Since the matrices which appear in (135) are of rank one and they are not a linear combination of each other, (135) is indefinite in general.

In this case improvement conditions for both estimators are more informative. Considering (84) and (44) these are

$$x'Nx \leqslant 1 \quad \text{for BLUE} \tag{136}$$

$$x' (N^{-1} + 2x_0x_0')^{-1} x \leqslant 1 \qquad \text{for BLE} \tag{137}$$

Since the difference (136)–(137) is larger than zero for $x_0 \neq 0$ as a result of Theorem A.7, the improvement condition for BLE is comparatively easier to achieve than the improvement condition for BLUE.

Let us now consider a general property of shrinkage estimators which can also be observed in BLE and BLUE. These estimators can be expressed in the following forms respectively

$$\hat{x}_1 = d_1 \cdot x_0 \tag{138}$$

$$\hat{x}_2 = d_2 \cdot x_0 \tag{139}$$

where

$$d_1 := \frac{x_0' N \hat{x}_s}{1 + x_0' N x_0} \tag{140}$$

$$d_2 := \frac{x_0' N \hat{x}_s}{x_0' N x_0} \tag{141}$$

and $\hat{x}_s$ is the GLSE of x, x_0 is the prior information vector about x as previously defined. Since d_1 and d_2 are scalar quantities which operate on x_0 in the above representations, the usefulness of BLE and BLUE depends closely on the relative magnitudes and directions of the components of x_0 with respect to the components of the sample estimates $\hat{x}_s$ (or equivalently x). If the elements of x_0 are appropriate on these a priori grounds, BLE and BLUE give reasonable estimates. On the other hand, if the elements of x_0 are different in magnitude and directions with respect to the elements of $\hat{x}_s$, the null hypothesis testing procedure given in section 4.2 can easily detect these inconsistencies. This can be seen directly from the test statistics given by 4.2.10 as

$$\gamma := (x_0 - \hat{x}_s)'N (x_0 - \hat{x}_s) \sim \chi^2{}_m \tag{142}$$

70

γ will be larger when x_0 is different than $\hat{x}_s$ in direction and magnitude. In this case it is dangerous to use these estimators except when deformations are below the noise level of observations as implied by the improvement conditions derived for BLE and BLUE. If these conditions hold, the direction and the magnitude of x_0 are no longer relevant because $\hat{x}_s$ in (138) and (139) will shrink the estimates toward zero.

Meanwhile improvements over GLSE can also be obtained if the prior information about the parameter vector to be estimated is stochastic in nature. In this case BLUUE can be used to introduce this information. This technique corresponds to the well-known "least square solution in the case of observation equations with weighted parameters" (Uotila, 1980). In section 4.6 this technique was examined within the scope of the proposed algorithmic approach. It was shown that introduction of prior information improves the results with respect to GLSE. The condition of improvement under incompatible prior information was given. Results are due to the discussion paper of Terasvirta (1979), except Corollary 2 which is about the amount of total bias committed by the use of incompatible prior information, is new.

Corollary 1 provides additional insights about the properties and usefulness of this estimator. From (116) the improvement condition reads as

$$ s'\ (\Sigma_{ee} + N^{-1})^{-1}s \leqslant 1 \tag{143} $$

the presence of s in the above condition indicates that the magnitude of systematic error vector (s) is important for (143) to hold rather than the amount of true parameter vector x in the case of BLE and BLUE. Therefore, this estimator is not robust against incompatible prior information. The stochastic nature of prior information, however, compensates for this disadvantage of BLUUE as we shall demonstrate using the following simplifying assumptions which also provide further insight about the properties of BLUUE under incompatible prior information.

71

If

$$u \sim (0, \sigma_u^2 I) \qquad\qquad (144)$$

$$e \sim (0, \sigma_e^2 I) \qquad\qquad (145)$$

$$k := \sigma_u^2 / \sigma_e^2 \qquad\qquad (146)$$

where σ_u^2 and σ_e^2 are the a priori variances of the disturbances and the additional information respectively, the (143) reduces to

$$\sigma_u^{-2} \; s' \; [k^{-1}I + (A'A)^{-1}]^{-1} s \leqslant 1 \qquad\qquad (147)$$

Using Theorem A.5, $(A'A) = C\Lambda C'$, (147) can be written as

$$\sum_{i=1}^{m} \frac{t_i^2 / \sigma_u^2}{k^{-1} + \lambda_i} \leqslant 1 \qquad\qquad (148)$$

where t_i are elements of $t = C's$ and λ_i are the eigenvalues of $(A'A)^{-1}$. Therefore, (143) under the assumptions given by (144) and (145), is likely to hold when

(i)   Bias-to-noise ratio $(s's/\sigma_u^2)$ is sufficiently small

(ii) λ_i are large

(iii) x_{0i} is not erroneous (i.e., s_i is not too large) in directions where there is little information (i.e., when λ_i is small)

(iv) k^{-1} is large (i.e., when σ_u^2 is sufficiently small or σ_e^2 is large.

Due to the stochastic nature of prior information a direct analytical comparison of BLUUE with BLE and BLUE which use prior information in a nonstochastic sense is not meaningful. Conditions such as (130) and (143) allow the experimenter to make qualitative inferences about these estimators. The following upper bounds for (130) which hold both for BLE and BLUE, and (143) which holds for BLUUE can easily be proven. Considering the minimum eigenvalues of $(A'A)^{-1}$ these are

$$\frac{1}{\lambda_{min}} \; \frac{x'x}{\sigma_u^2} \; < \; 1 \qquad\qquad (149)$$

$$\frac{1}{\lambda_{min}} \; \frac{s's}{\sigma_u^2} \; < \; 1 \qquad\qquad (150)$$

72

where λ_{min} is the minimum eigenvalue of $(A'A)^{-1}$.

Since λ_{min} and $\sigma_u{}^2$ is the same in both expressions the preference of one estimator over the other depends on the other terms. If the magnitude of the parameter vector x is expected to be small enough (in the scope of crustal movement analysis this corresponds to very small deformations) with respect to the magnitude of possible biases (s) in x_0, BLE and BLUE are preferable over BLUUE. However, if there exist pronounced deformations and the observations are precise enough (149) cannot be fulfilled. Therefore the usefulness of these two estimators are limited to the amount of signals.

On the other hand, the advantage of BLUUE over the other estimators lies in the fact that the experimenter can control the effect of systematic errors by the gradual reduction of its effect through its uncertainty σ_u^2 in $\tilde{x}_3$. In addition, the above expressions allow the experimenter to achieve this efficiency without extensive simulation studies.

The advantages of using stochastic prior information can further be exploited by the introduction of the concept of random parameters. This interpretation opens the possibility of employing another group of estimators which use prior information about the expected values of stochastic deformation parameters. This is the subject of the following chapter.

Chapter 5
PREDICTION USING PRIOR INFORMATION

5.1 INTRODUCTION

In our consideration of the improved estimation problem in the previous chapter, we have assumed that the deformation parameters to be estimated were fixed (deterministic) though prior information available about them was stochastic or fixed in nature.

The interpretation and availability of stochastic prior information as well as its possible discrepancies with the information obtained from sample observations (for instance, the difference between relative plate velocities implied by the long-term average models of global tectonics and the ones which are deduced from the current local and regional strain accumulation measurements for the Pacific and North American plates indicated by (Lyzanza et al., 1985)) suggest that these parameters themselves may act as random variables which can take on different values at different places and at different times. Within this interpretation, if a realistic density function can be postulated for these random variables, the problem of improved estimation in the presence of prior information, which is in this case about the expected values of the parameters, can also be examined using the Bayesian inference methods. In other words, prior information with its prior probability combined with a likelihood function using Bayes' Theorem yield the estimates of deformation parameters and their posterior probabilities.

Existing knowledge about the nature of crustal movements, unfortunately, is not sufficient (at least for today) to postulate a realistic density function about these deformation parameters. Nevertheless, the estimation problem with prior information can also be formulated independent of the density function of random deformation

74

parameters using the sampling theoretic approach.

In this framework two approaches are identified. The conventional approach which estimates the mean (expected values) of stochastic parameters and the estimation of random variables themselves which may be relevant when the short-term realization of random deformations is of interest. Estimation of stochastic parameters is also called prediction (Rao, 1965; Schaffrin, 1983).

In this chapter we formulate the improved estimation problem as follows. Deformation parameters are stochastic in nature. They exhibit a temporal and spatial behavior which fluctuates about an a priori known fixed mean and an a priori known dispersion. Although stochastic deformation parameters cannot be observed directly, they can be realized from another random vector (such as baseline differences) which can be obtained through observations (repeated baseline measurements). The estimation of random variables themselves combined with the prior information about their means is under consideration.

Three prediction techniques, namely: Best Homogeneously Linear Prediction (HOMBLIP), Best Homogeneously Unbiased Prediction (HOMBLUP), and Best Inhomogeneously Unbiased Prediction (INHOMBLUP) which are proposed by Schaffrin (1983) for the estimation of random deformation parameters in the presence of prior information, are examined in this chapter.

Since the purpose of introducing prior information in the proposed approach to crustal deformation analysis is to improve the estimation of random parameters, they are compared against a reference estimator which uses only observations. This predictor, as it is demonstrated in section 5.4, turns out to be GLSE. On the other hand, prior information about the expected values of random parameters are not known in practice with certainty. Therefore the effect of noncompatible prior information needs to be examined as we did in the previous chapter for the functional relationship linear model.

Although the assumption of random deformation parameters seems to be intuitively plausible in crustal movement analysis, especially for the plate tectonics models where additional information about the plate

75

motion parameters are derived from data which spans over geological time intervals, it is also possible to check the validity of the following model using the principle of the likelihood ratio developed by Rao (1965).

In this chapter the following linear model is under consideration

$$y = Ax + u \qquad (1a)$$

where y is an n×1 vector of random variables, A is an n×m (m<n) fixed design matrix of full rank. u is an n×1 vector of disturbances which are random and unobservable. x is the m×1 vector of unknown stochastic parameters. u and x have the following distributional properties

$$u \sim (0, \Sigma_{uu}), \qquad \Sigma_{uu} > 0 \qquad (1b)$$

$$x \sim (\mu, \Sigma_{xx}), \qquad \Sigma_{xx} > 0 \qquad (1c)$$

$$E(ux') = 0 \qquad (1d)$$

where Σ_{uu} is the n×n covariance matrix of disturbances. μ and Σ_{xx} are the m×1 vector of the expected values of x and m×m covariance matrix of the random (stochastic) parameter vector x respectively. It is assumed that the random parameter vector x and disturbances u are statistically independent. As a consequence of (1a)-(1d), the following relationships hold

$$E(y) = A\mu \qquad (2a)$$

$$E[y - E(y)][y - E(y)]' = A\Sigma_{xx}A' + \Sigma_{uu} \qquad (2b)$$

$$E[x - E(x)][y - E(y)]' = \Sigma_{xx}A' \qquad (2c)$$

Under this setup, (1)-(2), we shall consider the problem of estimating (predicting) the random variable x in the presence of prior information about μ and Σ_{xx}.

5.2 COMPATIBILITY TESTING

If the model parameters are interpreted as random variables with certain mean and variances available to the experimenter as a result of prior investigations, then this information and the information implied by the sample observations can be tested, as in the case of deterministic

76

parameters, to decide whether they are compatible. This step can be used to check the validity of prior information about the expected values of random parameters before they are introduced into the improved estimation procedure within the scope of the proposed approach. In this section a null hypothesis testing procedure is developed for this purpose.

Consider the linear model given by (1)-(2). Let μ_0 denote the known prior information about the expected values of random parameters and Σ_{xx} is the known positive definite m×m covariance matrix of x. Then the null hypothesis H_0 is

$$H_0: \quad \mu_0 = \mu := E(x) \tag{3}$$

<u>Lemma 1</u>: (Rao, 1965) Let $\hat{\mu} := Gy + d$ be an inhomogeneous linear estimator of $\mu := E(x)$ in the linear model (1)-(2). Then the optimum values of d, G for which

i) $E(\mu - \hat{\mu}) = 0$ for all values of μ $\hspace{2cm}$ (4a)

ii) $E(\mu - \hat{\mu})(\mu - \hat{\mu})'$ is a minimum $\hspace{1.5cm}$ (4b)

are

$$d = 0 \tag{5a}$$
$$G = N^{-1} A' \Sigma_{uu}^{-1} y \tag{5b}$$

which imply

$$\hat{\mu} := Gy + d = N^{-1} A' \Sigma_{uu}^{-1} y \tag{5c}$$

The dispersion matrix of $\hat{\mu}$ is given by

$$D(\hat{\mu}) = (\Sigma_{xx} + N^{-1}) \tag{5d}$$

where $N := A' \Sigma_{uu}^{-1} A$.

In addition, if $x \sim N(A\mu, \Sigma_{xx})$, $u \sim N(0, \Sigma_{uu})$ and x and u are stochastically independent, as implied by (2), then

i) $\hat{\mu} = N^{-1} A' \Sigma_{uu}^{-1} y \sim N(\mu, \Sigma_{xx} + N^{-1})$ (6a)

ii) $\hat{\mu}$ is inference sufficient for the parameter μ (6b)

Therefore the best inhomogeneously linear estimator $\hat{\mu}$ of μ with respect to the conditions (4a)-(4b) is the same as GLSE. But the dispersion matrix of the estimator is different (note that this estimator is independent of μ).

<u>Proposition 1</u>: The following quadratic form is χ^2 distributed with m degrees of freedom as a result of Lemma 1

$$(\mu - \hat{\mu})' (\Sigma_{xx} + N^{-1})^{-1} (\mu - \hat{\mu}) \sim \chi_m^2$$ (7a)

<u>Proof</u>: Let

$$\delta := \mu - \hat{\mu}$$ (7b)

then from (5c) and (1a)

$$\delta = \mu - x - N^{-1} A' \Sigma_{uu}^{-1} u$$ (7c)

It is easy to show that

$$\delta \sim (0, \Sigma_{xx} + N^{-1})$$ (7d)

Since $\Sigma_{xx} + N^{-1} > 0$, it can be expressed as (Theorem A.2)

$$\Sigma_{xx} + N^{-1} = Z^{\frac{1}{2}} Z^{\frac{1}{2}}$$ (7e)

where

$$Z^{\frac{1}{2}} = (\Sigma_{xx} + N^{-1})^{\frac{1}{2}}$$ (8a)

and

$$Z^{-\frac{1}{2}} \delta \sim (0, I)$$ (8b)

Using Theorem A.11

$$Z^{-\frac{1}{2}} \delta \sim N(0, I)$$ (8c)

78

and as a result of Theorem A.10

$$\delta' Z^{-1} \delta = (\mu - \hat{\mu})' \; (\Sigma_{xx} + N^{-1})^{-1} \; (\mu - \hat{\mu}) \sim \chi_m^2 \qquad (8d)$$

This completes the proof.

Now substituting μ_0 for μ in (8d) we obtain a test statistic to check the validity of prior information μ_0 about the expected values of x, before it is introduced into the estimation process. If α is a predefined error probability of the first kind, then the null hypothesis H_0 will be accepted for $\delta' Z^{-1} \delta \leqslant \chi_{m, 1-\alpha}^2$. Note the interesting structural duality with the testing procedure in section 4.2 of the functional relationships model.

5.3 COMPARISON CRITERION

The statistical theory used in the previous section for comparing different estimators can also be extended to cover the linear prediction problems.

The comparison criterion is formulated as follows.

<u>Definition 1</u>: Let x be an m × 1 vector of random variables such that $E(x) =: \mu$ and let $\hat{x}$ denote the corresponding predicted values. The matrix valued Mean Square Error of Prediction (MSEP) of $\hat{x}$ is then defined by

$$MSEP(\hat{x}) := E(x - \hat{x})(x - \hat{x})' = D(x - \hat{x}) + bb' \qquad (9)$$

where b is the bias vector of prediction and is defined as

$$b := E(x - \hat{x}) \qquad (9a)$$

(Schaffrin, 1983). Now the comparison criterion is formulated.

<u>Definition 2</u> (MSEP Criterion): A predictor $\hat{x}_1$ of the random variable x is said to be better than the another predictor $\hat{x}_2$ of x if the difference matrix

79

$$\text{MSEP}(\hat{x}_2) - \text{MSEP}(\hat{x}_1) \tag{10}$$

is positive semi definite (p.s.d.), where

$$\text{MSEP}(\hat{x}_1) = E(x - \hat{x}_1)(x - \hat{x}_1)' \tag{11}$$

$$\text{MSEP}(\hat{x}_2) = E(x - \hat{x}_2)(x - \hat{x}_2)'. \tag{12}$$

Consider now the following case. Let $\hat{x}_s$ be an unbiased predictor of x, and furthermore let this predictor employ only the sample data. Let $\hat{x}$ be another predictor of x, possibly biased. If Definition 2 holds, then

$$b'b \leqslant \text{tr}[D(x - \hat{x}_s) - D(x - \hat{x})] \tag{13}$$

where b is the bias vector of $\hat{x}$. This property is similar to (4.19) and forms the logical basis of using MSEP criterion. Now, as in Section 4.3, it is necessary to define a reference predictor. This is the topic of the following section.

5.4 PREDICTION USING SAMPLE INFORMATION ONLY

Consider the linear model described by (1) - (2) with the exception that no prior information is available about the random parameters. The following theorem defines the sample predictor.

<u>Theorem 1:</u> Let Gy be a linear predictor of the random parameters x in the linear model described by (1) - (2). Then the optimum value of G (this value is denoted by the same letter for the sake of simplicity in the notation) for which

$$E(x - Gy)'(x - Gy) = \min_{G} \tag{14}$$

subject to

$$E(x - Gy) = (I - GA)\mu = 0 \tag{15a}$$

for all μ, that is,

$$I - GA = 0 \tag{15b}$$

is

$$G = (A'\Sigma_{uu}^{-1}A)^{-1}A'\Sigma_{uu}^{-1}$$
$$\hat{x}_s := Gy = (A'\Sigma_{uu}^{-1}A)^{-1}A'\Sigma_{uu}^{-1}y. \tag{16}$$

The dispersion matrix of the predictor $\hat{x}_s$ is

$$D(\hat{x}_s) = D(x - \hat{x}_s) = (A'\Sigma_{uu}^{-1}A)^{-1}. \tag{17}$$

<u>Proof</u>: Considering (14), let the target function to be minimized be

$$Q := \operatorname{tr}[E(x - Gy)(x - Gy)' + 2\Lambda'(I - GA)'] \tag{18}$$

where Λ is a Lagrange multiplier matrix. Substituting (14) in (18) we obtain

$$Q = \operatorname{tr}[G\Sigma_{uu}G' + 2\Lambda'(I - GA)']. \tag{19}$$

Now minimizing Q with respect to G

$$G\Sigma_{uu} - \Lambda'A' = 0 \tag{20}$$

$$G = \Lambda'A'\Sigma_{uu}^{-1} \tag{21}$$

and multiplying both sides of (21) by A on the right-hand side and considering (15)

$$I = \Lambda'A'\Sigma_{uu}^{-1}A \tag{22}$$

$$\Lambda' = (A'\Sigma_{uu}^{-1}A)^{-1} \tag{23}$$

is obtained. From (21) it follows that

$$G = (A'\Sigma_{uu}^{-1}A)^{-1}A'\Sigma_{uu}^{-1} \tag{24}$$

and

$$\hat{x}_s := Gy = (A'\Sigma_{uu}^{-1}A)^{-1}A'\Sigma_{uu}^{-1}y. \tag{25}$$

81

This completes the proof.

Observe that (25) is equivalent to GLSE. Hence, if no additional information is available, then the GLSE is also the best linear unbiased predictor in the sense of (15) and (19). This predictor can now be used as a reference predictor against which the effect of additional information that are used by alternative predictors can be compared.

5.5 BEST HOMOGENEOUSLY LINEAR PREDICTION WITH PRIOR INFORMATION

Theorem 1: Best Homogeneously Linear Prediction, HOMBLIP, (Schaffrin, 1983). Let Gy be a linear predictor of x in the linear model (1) - (2). Then the optimum value of G for which

$$E(x - Gy)'(x - Gy) = \min_{G} \tag{26}$$

yields to the HOMBLIP $\hat{x}_1$ of x

$$\hat{x}_1 := Gy = (I + \Sigma_{xx}N)^{-1}\Big\{\Sigma_{xx}A'\Sigma_{uu}^{-1}y + \big[(1 + \mu'\bar{N}\mu)^{-1}\mu'[(I + \Sigma_{xx}N)']^{-1} \cdot$$
$$A'\Sigma_{uu}^{-1}y] \cdot \mu\Big\} \tag{27}$$

where

$$N := A'\Sigma_{uu}^{-1}A \tag{28}$$

$$\bar{N} := N(I + \Sigma_{xx}N)^{-1} = (I + N\Sigma_{xx})^{-1}N \tag{29}$$

This predictor is biased and the resulting bias is given by

$$b := E(x - Gy) = (1 + \mu'\bar{N}\mu)^{-1}(I + \Sigma_{xx}N)^{-1}\mu \tag{30}$$

The mean square error matrix of the prediction is

$$MSEP(\hat{x}_1) = E(x - \hat{x}_1)(x - \hat{x}_1)'$$
$$= (I+\Sigma_{xx}N)^{-1}\Sigma_{xx} + (1+\mu'\bar{N}\mu)^{-1}(I+\Sigma_{xx}N)^{-1}\mu\mu'[(I+\Sigma_{xx}N)^{-1}]' \tag{31}$$

Since our interest is in the gain of using prior information, the following corollary is of interest.

Corollary 1: HOMBLIP $\hat{x}_1$ of x is better than the sample predictor $\hat{x}_s$ under Definition 2.

Proof: Let us again consider $MSEP(\hat{x}_1)$ and $MSEP(\hat{x}_s)$ given by (31) and (17). Then

$$MSEP(\hat{x}_s) - MSEP(\hat{x}_1) = \bar{N}^{-1}(I - \bar{N}\Sigma_{xx}) - (I - \Sigma_{xx}\bar{N})\Sigma_{xx}$$
$$- (1 + \mu'\bar{N}\mu)^{-1}(I - \Sigma_{xx}\bar{N}) \; \mu\mu'(I - \Sigma_{xx}\bar{N})' \quad (32)$$

Considering the following identity

$$\bar{N}^{-1} - \mu(1 + \mu'\bar{N}\mu)^{-1}\mu' \equiv \bar{N}^{-1}(\bar{N}^{-1} + \mu\mu')^{-1}\bar{N}^{-1} \quad (33)$$

(32) can be written as

$$MSEP(\hat{x}_s) - MSEP(\hat{x}_1) = (I - \Sigma_{xx}\bar{N})\bar{N}^{-1}(\bar{N}^{-1} + \mu\mu')^{-1}\bar{N}^{-1}(I - \Sigma_{xx}\bar{N})' \quad (34)$$

Since all the matrices within the parentheses are positive definite, (32) is also positive definite. This completes the proof.

Consider now the HOMBLIP $\hat{x}_1$ of x rewritten in the following modified form

$$\hat{x}_1 = [B\Sigma_{xx} + B\mu(1 + \mu'\bar{N}\mu)^{-1}\mu'B']A'\Sigma_{uu}^{-1}y \quad (35)$$

where

$$B := (I - \Sigma_{xx}\bar{N}) \equiv (I + \Sigma_{xx}N)^{-1} \quad (36)$$

As in the case of BLE, BLUE and BLUUE, this predictor is practical if the expected value of the stochastic parameter and its covariance matrix are both available as prior information. To give a more realistic validity to the prediction of crustal motion parameters, it would be logical to assume that this prior information is of the form μ_0, which is an m×1 nonzero vector, and that it is different than the true value μ, i.e.,

83

$$\mu_0 \neq \mu := E(x) \tag{37}$$

Such deviations are quite likely to hold in crustal movement analysis; for instance, due to the inhomogeneous properties of the earth's crust where the prior information is not fully representative of the long term averages at different areas. Another possibility for such deviations could be some systematic model or measurement errors which occur in the process of obtaining the prior information. However, it is likely that this information may contain a component of reality which can be used to improve the prediction of the random parameters in the sense of Definition 2 under controlled circumstances. These conditions are the topic of the following discussion.

Consider the bias vector of the prediction when $\mu \neq \mu_0$. From (9a)

$$b_0 := E(x - \hat{x}_1) = (I - G_0 A)\mu \tag{38}$$

where

$$G_0 = [B\Sigma_{xx} + B\mu_0(1 + \mu_0'\bar{N}\mu_0)^{-1}\mu_0'B']A'\Sigma_{uu}^{-1} \tag{39}$$

and using identity (33) the following expression for the bias vector is obtained

$$b_0 = B\bar{N}^{-1}(\bar{N}^{-1} + \mu_0\mu_0')^{-1}\mu \tag{40}$$

Let us now investigate the improvement conditions. When $\mu_0 \neq \mu$ the predictor (27), which will now be denoted by $\tilde{x}_1$, has the following form

$$\tilde{x}_1 := G_0 y = (I + \Sigma_{xx}N)^{-1}\Big\{\Sigma_{xx}A'\Sigma_{uu}^{-1}y + \big[(1 + \mu_0'\bar{N}\mu_0)^{-1}\mu_0'[(I + \Sigma_{xx}N)']^{-1}$$
$$\cdot A'\Sigma_{uu}^{-1}y] \cdot \mu_0\Big\} \tag{41}$$

and its corresponding MSEP is

$$MSEP(\tilde{x}_1) := E(x - \tilde{x}_1)(x - \tilde{x}_1)'$$
$$= B\Sigma_{xx} + B\mu_0(1 + \mu_0'\bar{N}\mu_0)^{-2}(\mu_0'\bar{N}\mu_0)\mu_0'B' + b_0 b_0' \tag{42}$$

Considering the MSEP of the sample predictor $\hat{x}_s$

$$MSEP(\hat{x}_s) = D(x - \hat{x}_s) = D(\hat{x}_s) = N^{-1} \equiv \bar{N}^{-1}B \qquad (43)$$

the corresponding difference matrix, under the general condition of betterness, Definition 2, is

$$MSEP(\hat{x}_s) - MSEP(\hat{x}_1) = B\Big\{\bar{N}^{-1} - \mu_0(1 + \mu_0'\bar{N}\mu_0)^{-2}(\mu_0'\bar{N}\mu_0)\mu_0'$$
$$- [I - \mu_0(1 + \mu_0'\bar{N}\mu_0)^{-1}\mu_0'\bar{N})]\mu\mu'$$
$$\cdot [I - \mu_0(1 + \mu_0'\bar{N}\mu_0)^{-1}\mu_0'\bar{N}]'\Big\}B' \qquad (44)$$

This expression is p.s.d. if the matrix differences within the curly brackets is p.s.d. Using Theorem A.1 and considering

$$\big[\bar{N}^{-1} - \mu_0(1 + \mu_0'\bar{N}\mu_0)^{-2}(\mu_0'\bar{N}\mu_0)\mu_0'\big] \geqslant 0 \qquad (45)$$

for all μ_0, it can be shown that (44) is p.s.d. iff

$$\mu'\bar{N}\Big[\bar{N}^{-1} - \frac{\mu_0\mu_0'}{\frac{1}{2} + \mu_0'N\mu_0}\Big]\bar{N}\mu \leqslant 1 \qquad (46)$$

This expression can also be written, using Theorem A.9, as

$$\mu'\big(\bar{N}^{-1} + 2\mu_0\mu_0'\big)^{-1}\mu \leqslant 1 \qquad (47)$$

This result proves the following theorem.

Theorem 2:   The necessary and sufficient condition for the biased predictor $\tilde{x}_1$ to be better than the sample predictor $\hat{x}_s$ for all $\mu_0 \neq \mu$ under Definition 2 is

$$\mu'(\bar{N}^{-1} + 2\mu_0\mu_0')^{-1}\mu \leqslant 1 \qquad (48)$$

Now observe that

$$\bar{N} - (\bar{N}^{-1} + 2\mu_0\mu_0')^{-1} \geqslant 0 \qquad (49a)$$

for all μ_0, as a result of Theorem A.7. Therefore,

$$\mu' \bar{N} \mu \leqslant 1 \qquad (49b)$$

is an upper bound for (48). If now, $u \sim (0, \sigma_u^2 I)$, and $x \sim (u, \sigma_u^2/k \cdot I)$ where k is a proportion constant relating the variances of u (σ_u^2) and x (σ_x^2), then (49) can be written as

$$\sigma_u^{-2} \mu' [k^{-1} I + (A'A)^{-1}]^{-1} \mu \leqslant 1 \qquad (50)$$

This expression is an increasing function of k and it approaches zero in the limit

$$\lim_{k \to 0+} \{ \sigma_u^{-2} \mu' [k^{-1} I + (A'A)^{-1}] \mu \} = 0 \qquad (51)$$

Therefore the following corollary holds.

Corollary 2: If $u \sim (0, \sigma_u^2 I)$ and $x \sim (u, \sigma_u^2/k \cdot I)$, then a k^* always exists such that $MSEP(\hat{x}_s) - MSEP(\tilde{x}_1) \geqslant 0$ for all $\mu_0 \neq \mu$ where $0 \leqslant k \leqslant k^*$. Now let

$$z := \bar{N}^{\frac{1}{2}} \mu \qquad (52)$$
$$z_0 := \bar{N}^{\frac{1}{2}} \mu_0 \qquad (53)$$

and consider the identity

$$B := (I - \Sigma_{xx} \bar{N}) \equiv (I + \Sigma_{xx} N)^{-1} \qquad (54)$$

then, the quadratic bias from (40) and the condition of improvement from Theorem 2 each take the following forms respectively

$$b_0' b_0 = z' (I + z_0 z_0')^{-1} \bar{N}^{-\frac{1}{2}} BB' \bar{N}^{-\frac{1}{2}} (I + z_0 z_0')^{-1} z \qquad (55)$$

$$z' (I + 2 z_0 z_0')^{-1} z \leqslant 1$$

It is now easy to prove the following corollary using the same steps followed in Corollary 2 of BLE.

Corollary 3: The quadratic bias due to the erroneous prior information is bounded under the condition of improvement and

86

$$\sup \; b_0 b_0' \; \leqslant \; \lambda_m \tag{56}$$

$$z'(I+2z_0z_0')^{-1} \; \leqslant \; 1$$

where λ_m is the largest eigenvalue of $(I + N\Sigma_{xx})^{-1}N^{-1}$.

5.6 BEST HOMOGENEOUSLY LINEAR UNBIASED PREDICTION WITH PRIOR INFORMATION

<u>Theorem 1</u>: Best Homogeneously Linear Unbiased Prediction, HOMBLUP, (Schaffrin, 1983). Let Gy be a linear predictor of x in the linear model (1) - (2), then the optimum value of G for which

$$E(x - Gy)'(x - Gy) = \min_{G} \tag{56}$$

subject to

$$E(x - Gy) = (I - GA)\mu = 0 \tag{57}$$

in the sense of

$$\mathrm{tr}\left[E(x - Gy)(x - Gy)' + 2(I - GA)\mu\lambda'\right] = \min_{G} \tag{58}$$

where λ' is a Lagrangian multiplier vector, leads to the HOMBLUP $\hat{x}_2$ of x

$$\hat{x}_2 := Gy$$
$$= (I + \Sigma_{xx}N)^{-1}\left\{(\Sigma_{xx}A'\Sigma_{uu}^{-1}y + \left[(\mu'\bar{N}\mu)^{-1}\mu'\left\{(I + \Sigma_{xx}N)^{-1}\right\}'A'\Sigma_{uu}^{-1}y\right]\cdot\mu\right\} \tag{59}$$

where

$$N := A'\Sigma_{uu}^{-1}A \tag{60}$$

$$\bar{N} := N(I + \Sigma_{xx}N)^{-1} \tag{61}$$

The mean square error matrix of the prediction is given by

$$\text{MSEP}(\hat{x}_2) = (I + \Sigma_{xx}N)^{-1}\mu(\mu'\bar{N}\mu)^{-1}\mu'[(I + \Sigma_{xx}N)^{-1}]' + (I + \Sigma_{xx}N)^{-1}\Sigma_{xx} \tag{62}$$

Note that this predictor, similar to BLUE (section 4.6), reduces to GLSE as a result of the weak unbiasedness condition (57) for the univariate case, thereby no improvements are possible through prior information.

Proposition 1: HOMBLUP $\hat{x}_2$ of x is better than the sample predictor $\hat{x}_s$ of x <u>with respect to Definition 2</u>.

<u>Proof</u>: Consider the difference of the MSEP matrices of $\hat{x}_2$ and $\hat{x}_s$ given by (62) and (17) respectively

$$\text{MSEP}(\hat{x}_s) - \text{MSEP}(\hat{x}_2) = N^{-1} - \Big\{ (I + \Sigma_{xx}N)^{-1}\mu(\mu'\bar{N}\mu)^{-1}\mu'[(I + \Sigma_{xx}N)^{-1}]'$$
$$+ (I + \Sigma_{xx}N)^{-1}\Sigma_{xx} \Big\} \tag{63}$$

which reduces after some manipulations to

$$\text{MSEP}(\hat{x}_s) - \text{MSEP}(\hat{x}_2) = (I - \Sigma_{xx}\bar{N})[\bar{N}^{-1} - \mu(\mu'\bar{N}\mu)^{-1}\mu'](I - \Sigma_{xx}\bar{N})' \tag{64}$$

Since $(I - \Sigma_{xx}\bar{N}) \equiv (I + \Sigma_{xx}N)^{-1}$ as a result of Identity A.9 and since the term within the brackets is symmetric, (64) is p.s.d. iff

$$[\bar{N} - \mu(\mu'\bar{N}\mu)^{-1}\mu'] \geqslant 0 \tag{65}$$

(Theorem A.6). Now let

$$B := \bar{N}^{-1} \tag{66}$$

$$d := \mu(\mu'\bar{N}\mu)^{-\frac{1}{2}} \tag{67}$$

then (65) can be written as B − dd′ and

$$B - dd' \geqslant 0$$

iff

$$d'B^{-1}d \leqslant 1$$

(Theorem A.1). Since, from (66) and (67)

$$d'B^{-1}d = (\mu'\bar{N}\mu)^{-1}(\mu'\bar{N}\mu) = 1$$

this implies

$$MSEP(\hat{x}_s) - MSEP(\hat{x}_2) \geqslant 0.$$

This completes the proof.

As shown in Corollary 1, this predictor is better than the sample predictor under the MSEP criterion of preference, and it is applicable if there exists a valid prior information μ_0 and a covariance Σ_{xx} such that

$$\mu_0 \propto \mu := E(x) \tag{68}$$

Yet this condition is seldom met in practice. Therefore, it is also necessary to examine the possibility of improvements when $\mu_0 \neq \mu$ (μ_0 is not proportional to μ). Consider the predictor given by (59). If the prior information μ_0 to be used in this expression (by replacing μ with μ_0 in (59)) is different than the expected value of the random parameters the resulting predictor is no longer unbiased. Let us denote this predictor under wrong prior information by $\tilde{x}_2$ then

$$\tilde{x}_2 = (I + \Sigma_{xx}N)^{-1}\left\{ (\Sigma_{xx}A'\Sigma_{uu}^{-1}y + [(\mu_0'\bar{N}\mu_0)^{-1}\mu_0'((I + \Sigma_{xx}N)^{-1})'A'\Sigma_{uu}^{-1}y]\cdot\mu_0 \right\} \tag{69}$$

where now $\mu_0 \neq \mu$ and the resulting bias can be obtained from

$$b_0 := E(x - \tilde{x}_2) = \mu - E(\tilde{x}_2) \tag{70}$$

Substitution of (69) into (70) gives the following bias vector

$$b_0 = (I - \Sigma_{xx}\bar{N})[I - \mu_0(\mu_0'\bar{N}\mu_0)^{-1}\mu_0'\bar{N}]\mu \tag{71}$$

Now, it is easy to see that for $\mu_0 = \mu$ bias vanishes. Similarily for $\mu_0 \propto \mu$ the prediction remains unbiased. Consequently, the incompatibility due

to scale differences between sample information and additional information on the expected values of the parameters do not have any influence on the predicted values.

To investigate the conditions under which biased prediction as a result of wrong prior information is better than the sample prediction consider the MSEP of the biased predictor $\tilde{x}_2$ of x with prior information $\mu_0 \neq \mu := E(x)$ and the nonvanishing bias vector b_0 given by (71). Then

$$MSEP(\tilde{x}_2) := E(x - \tilde{x}_2)(x - \tilde{x}_2)' \tag{72}$$

leads to

$$MSEP(\tilde{x}_2) = (I - \Sigma_{xx}\bar{N})\mu_0(\mu_0'\bar{N}\mu_0)^{-1}\mu_0'(I - \Sigma_{xx}\bar{N})' + (I - \Sigma_{xx}\bar{N})\Sigma_{xx} + b_0 b_0' \tag{73}$$

Let us also observe that

$$MSEP(\hat{x}_s) = D(\hat{x}_s) = N^{-1} \equiv \bar{N}^{-1}(I - \bar{N}\Sigma_{xx}) \tag{74}$$

Now, under the definition of improvement as given by Definition 2 it can be shown, after a simple transformation, that

$$MSEP(\hat{x}_s) - MSEP(\tilde{x}_2) = (I - \Sigma_{xx}\bar{N})[B - B\bar{N}\mu\mu'\bar{N}B](I - \Sigma_{xx}\bar{N})' \tag{75}$$

where

$$B := [\bar{N}^{-1} - \mu_0(\mu_0'\bar{N}\mu_0)^{-1}\mu_0'] \tag{76}$$

Since B is symmetric and p.s.d. for any μ_0, it can be expressed as $B = B^{1/2}B^{1/2}$ (Theorem A.2). Hence, (75) takes the following form

$$MSEP(\hat{x}_s) - MSEP(\tilde{x}_2) = C(I - dd')C' \tag{77}$$

where

$$C := (I - \Sigma_{xx}\bar{N})B^{1/2} \tag{78}$$

$$d := B^{1/2}\bar{N}\mu \tag{79}$$

90

Now, (77) is p.s.d. iff

$$d'd = \mu'\bar{N}\mu - \frac{\left(\mu'\bar{N}\mu_0\right)^2}{\mu_0'\bar{N}\mu_0} \leqslant 1 \tag{80}$$

as a result of Theorem A.1. However,

$$\sup_{\mu_0} \frac{(\mu'\bar{N}\mu_0)^2}{\mu_0'\bar{N}\mu_0} = \mu'\bar{N}\mu \implies d'd = 0 \tag{81}$$

(Theorem A.3). Also

$$\inf_{\mu_0} \frac{(\mu'\bar{N}\mu_0)^2}{\mu_0'\bar{N}\mu_0} = \lambda_0 \tag{82}$$

(Theorem A.4), where λ_0 is the minimum eigenvalue of

$$\bar{N}^{\frac{1}{2}}\mu\mu'\bar{N}^{\frac{1}{2}} \tag{83}$$

Since this matrix is of rank one, $\lambda_0 = 0$. Therefore, as a result (80), (75) is p.s.d. iff

$$d'd = \mu'\bar{N}\mu \leqslant 1 \tag{84}$$

Hence, the following theorem has been proven.

<u>Theorem 2</u>: A sufficient condition for the biased predictor $\tilde{x}_2$ of x to be better than the sample predictor $\hat{x}_S$ of x for all $\mu_0 \neq \mu := E(x)$ under the MSEP of preference is

$$\mu'N(I + \Sigma_{xx}N)^{-1}\mu \leqslant 1 \tag{85}$$

Now, if $u \sim (0, \sigma_u^2 I)$ and $x \sim (u, \sigma_u^2/k \cdot I)$ where k is a proportion constant relating the variances of u (σ_u^2) and x (σ_x^2), then the above condition can be written as

$$\sigma_u^{-2}\mu'[k^{-1}I + (A'A)^{-1}]^{-1}\mu \leqslant 1 \tag{86}$$

91

(note that this condition is independent of μ_0). This expression is an increasing function of k^{-1}, and it approaches zero as

$$\lim_{k \to 0+} \left\{ \sigma_u^{-2} \mu' \left[k^{-1}I + (A'A)^{-1} \right]^{-1} \mu \right\} = 0 \tag{87}$$

Therefore the following corollary holds.

Corollary 2: If $u \sim (0, \sigma_u^2 I)$ and $x \sim (u, \sigma_u^2/k \cdot I)$, then there always exists a k^* such that $MSEP(\hat{x}_s) - MSEP(\tilde{x}_2) > 0$ for all $\mu_0 \neq \mu$ where $0 < k < k^*$.

Now let

$$z := \bar{N}^{\frac{1}{2}} \mu \tag{88}$$

$$z_0 := \bar{N}^{\frac{1}{2}} \mu_0 \tag{89}$$

then the quadratic bias, from (71), is given by

$$b_0' b_0 = z'(I - z_0(z_0' z_0)^{-1} z_0') \bar{N}^{-\frac{1}{2}} (I - \Sigma_{xx} \bar{N})' (I - \Sigma_{xx} \bar{N}) \bar{N}^{-\frac{1}{2}}$$

$$\cdot (I - z_0(z_0' z_0)^{-1} z_0') z \tag{90}$$

and the condition of improvement under erroneous prior information (85) transforms to

$$z'z < 1 \tag{91}$$

The following corollary holds.

Corollary 3: The quadratic bias due to the erroneous prior information is bounded under the condition of improvement and

$$\sup_{z'z < 1} b_0' b_0 < \lambda_m \tag{92}$$

where λ_m is the largest eigenvalue of $(I + N\Sigma_{xx})^{-1} N^{-1}$.

<u>Proof</u>: Considering

$$\sup_{z_0} z'[I - z_0(z_0')^{-1}z_0']z = z'z \leqslant 1 \qquad (93)$$

and following Definition A.4

$$\sup_{\substack{z'z \leqslant 1}} b_0'b_0 \leqslant \lambda_m \qquad (94)$$

where λ_m is the largest eigenvalue of

$$\bar{N}^{-\frac{1}{2}}(I - \bar{N}\Sigma_{xx})(I - \bar{N}\Sigma_{xx})'\bar{N}^{-\frac{1}{2}} \qquad (95)$$

Now, using Theorem A.9, it can be shown that

$$\text{tr}[\bar{N}^{-\frac{1}{2}}(I - \bar{N}\Sigma_{xx})(I - \bar{N}\Sigma_{xx})'\bar{N}^{-\frac{1}{2}}] = \text{tr}[(I - \bar{N}\Sigma_{xx})(I - \bar{N}\Sigma_{xx})'\bar{N}^{-1}]$$

$$= \text{tr}[(I + N\Sigma_{xx})^{-1}N^{-1}] \qquad (96)$$

Therefore the largest eigenvalue of (95) is equal to the largest eigenvalue of $(I + N\Sigma_{xx})^{-1}N^{-1}$. This completes the proof.

5.7 BEST INHOMOGENEOUSLY LINEAR UNBIASED PREDICTION WITH PRIOR INFORMATION

<u>Theorem 1</u>: Best Inhomogeneously Linear Unbiased Prediction, INHOMBLUP, (Schaffrin, 1983). Let Gy + d be an inhomogeneous linear predictor of x in the linear model given by (1) - (2). Then the optimum value of G and d for which

$$E(x - Gy - d)'(x - Gy - d) = \min_{G,d} \qquad (97)$$

subject to

$$b := E(x - Gy - d) = (I - GA)\mu - d = 0 \qquad (98)$$

leads to the INHOMBLUP $\hat{x}_3$ of x, which is given by

$$\hat{x}_3 = (I + \Sigma_{xx}N)^{-1}(\Sigma_{xx}A'\Sigma_{uu}^{-1}y + \mu) \tag{99}$$

where, $N := A'\Sigma_{uu}^{-1}A$. The mean square error matrix of the prediction is

$$MSEP(\hat{x}_3) = D(x - \hat{x}_3) = (N + \Sigma_{xx}^{-1})^{-1} \tag{100}$$

<u>Corollary 1</u>:   INHOMBLUP $\hat{x}_3$ of x is a better predictor than the sample predictor $\hat{x}_s$ of x according to Definition 2.

<u>Proof</u>:   Consider the $MSEP(\hat{x}_3)$, from (100), and $MSEP(\hat{x}_s)$, from (17), then

$$MSEP(\hat{x}_s) - MSEP(\hat{x}_3) = N^{-1} - (N + \Sigma_{xx}^{-1})^{-1} \geqslant 0 \tag{101}$$

following Theorem A.7.

Again, as in the previous cases, for this predictor to be applicable additional information about $E(x) =: \mu$ and its covariance matrix Σ_{xx} needs to be available. However if this information μ_0 is not compatible, that is $\mu_0 \neq \mu$, then the prediction is no longer unbiased (<u>this biased predictor is now denoted by $\tilde{x}_3$</u>) and the bias of the prediction is given by

$$b_0 := E(x - \tilde{x}_3) = (I + \Sigma_{xx}N)^{-1}(\mu - \mu_0) \tag{102}$$

and the biased predictor is

$$\tilde{x}_3 = (I + \Sigma_{xx}N)^{-1}(\Sigma_{xx}A'\Sigma_{uu}^{-1}y + \mu_0) \tag{103}$$

Considering (102) and (103) the following theorem holds.

<u>Theorem 2</u>:   The necessary and sufficient condition for the biased predictor $\tilde{x}_3$ of x to be better than the sample predictor $\hat{x}_s$ of x for all $\mu_0 \neq \mu$ according to Definition 2 is

$$(\mu - \mu_0)'\bar{N}(\mu - \mu_0) \leqslant 1 \tag{104}$$

<u>Proof</u>:   The mean square matrix of the biased predictor $\tilde{x}_3$ of x is, from

$$MSEP(\tilde{x}_3) = D(\tilde{x}_3 - x) + b_0 b_0' \tag{105}$$

94

where b_0 is given by (102) and

$$D(x - \tilde{x}_3) = (N + \Sigma_{xx}^{-1})^{-1} \tag{106}$$

Now the difference matrix implied by Definition 2 is

$$MSEP(\hat{x}_s) - MSEP(\tilde{x}_3) = N^{-1} - \left[(N + \Sigma_{xx}^{-1})^{-1} + b_0 b_0'\right] \tag{107}$$

which reduces after some manipulations to

$$MSEP(\hat{x}_s) - MSEP(\tilde{x}_3) = B(\bar{N}^{-1} - ss')B' \tag{108}$$

where

$$B := (I + \Sigma_{xx}N)^{-1} \tag{109}$$

$$s := \mu - \mu_0 \tag{110}$$

Therefore, (108) is p.s.d. iff $s'\bar{N}s \leqslant 1$ (Theorem A.1). This completes the proof.

Now assume that

$$u \sim (0, \sigma_u^2 I) \tag{111a}$$

$$x \sim (\mu, \sigma_u^2/k \cdot I) \tag{111b}$$

where k is a proportion constant. Then (104) takes the following form

$$s'\bar{N}s = s'\left[k^{-1}I + (A'A)^{-1}\right]^{-1}s \leqslant 1 \tag{112}$$

This expression is again an increasing function of k and it approaches zero in the limit. Thus the following corollary holds.

Corollary 2: If $u \sim (0, \sigma_u^2 I)$ and $x \sim (\mu, \sigma_u^2/k \cdot I)$ then there always exists a k^* such that $MSEP(\hat{x}_s) - MSEP(\tilde{x}_3) \geqslant 0$ for all $\mu_0 \neq \mu$ where $0 \leqslant k \leqslant k^*$.

Let us now examine the amount of bias introduced by the use of

incompatible prior information, under the improvement condition given by (104).

Corollary 3: The total bias due to the erroneous prior information under the condition of improvement (104) is bounded. The supremum of it is equal to or less than the largest eigenvalue λ_m of $(I + N\Sigma_{xx})^{-1}N^{-1}$, i.e.,

$$\sup_{x_0} b_0' b_0 \leq \lambda_m \tag{113}$$

$$\text{s.t.:} \quad s'\bar{N}s \leq 1$$

The proof follows the similar lines given in Corollary 4.6.2.

5.8 FURTHER DISCUSSIONS

In the previous sections three different predictors were examined. Let us emphasize that prediction in technical or in everyday usage has different connotations. It can be defined generally as a statement about an unknown and uncertain event. Prediction in this study implies the estimation of stochastic parameters whose meaning is also encapsulated by the preceding general definition (for a comprehensive discussion, confer (Bibby and Tautenburg, 1977)). In the linear model described by (1) one might be interested in the estimation of the expected values of random parameters (classical approach) or the estimation of random parameters themselves (Rao, 1965; Schaffrin, 1983). Within the scope of crustal movement analysis the latter implies that short-term deformations which are regarded as stochastic with certain expectations over longer periods of time are of interest. Within this context, the above examined prediction techniques estimate short-term realization of random deformation parameters provided that prior information about their long-term behavior (their expected values and dispersion matrix) are known beforehand.

It was demonstrated in section 5.4 that the GLSE technique, in the absence of any prior information, is also a predictor which can be used

TABLE 2: PREDICTORS WITH PRIOR INFORMATION

	HOMBLIP	HOMBLIP	INHOMBLIP
ESTIMATOR $\mu_0 = \mu = E(x)$	$\hat{x}_1 = (I + \Sigma_{xx}N)^{-1}\{\Sigma_{xx}A'\Sigma_{uu}^{-1}y + [(1 + \mu_0'\bar{N}\mu_0)^{-1}\mu_0'((I + \Sigma_{xx}N)')^{-1}A'\Sigma_{uu}^{-1}y] \cdot \mu_0\}$	$\hat{x}_2 = (I + \Sigma_{xx}N)^{-1}\{\Sigma_{xx}A'\Sigma_{uu}^{-1}y + [((\mu_0'\bar{N}\mu_0)^{-1}\mu_0'((I + \Sigma_{xx}N)')^{-1}A'\Sigma_{uu}^{-1}y] \cdot \mu_0\}$	$\hat{x}_3 = (I + \Sigma_{xx}N)^{-1}(\Sigma_{xx}A'\Sigma_{uu}^{-1}y + \mu_0)$
IMPROVEMENT CONDITIONS WHEN μ_0 IS NOT COMPATIBLE	$\mu'(\bar{N}^{-1} + 2\mu_0\mu_0')^{-1}\mu \leqslant 1$ or $\mu'\bar{N}\mu \leqslant 1$	$\mu'\bar{N}\mu \leqslant 1$	$(\mu - \mu_0)'\bar{N}(\mu - \mu_0) \leqslant 1$
TOTAL BIAS INTRODUCED	$b_0'b_0 \leqslant \lambda_m\ \varepsilon\ (I + N\Sigma_{xx})^{-1}N^{-1}$	same as HOMBLIP	same as HOMBLIP
ASSUMPTIONS	$u \sim (0, \sigma_u^2 I),\ x \sim (\mu, \sigma_u^2/k \cdot I)$	same as HOMBLIP	same as HOMBLIP
CONTROL VARIABLES FOR IMPROVEMENTS	$\sum\limits_{i=1}^{m} \dfrac{t_i^2/\sigma_u^2}{k^{-1} + \lambda_i} \leqslant 1$ $(A'A)^{-1} = CAC'$, $\quad t = C'\mu$	same as HOMBLIP	$\sum\limits_{i=1}^{m} \dfrac{t_i^2/\sigma_u^2}{k^{-1} + \lambda_i} \leqslant 1$ $(A'A)^{-1} = CAC'$, $t = C'(\mu - \mu_0)$

$N: = A'\Sigma_{uu}^{-1}A$, $\quad \bar{N}: = N(I + \Sigma_{xx}N)^{-1}$

97

to estimate the random model parameters. Note that the expected values of random parameters can also be estimated using the estimator given in Lemma 1 which is similar to the GLSE but has a different dispersion matrix.

Since the purpose of using additional information is to improve the estimates, the above predictors were compared against the sample predictor GLSE. If the prior information is correct then its introduction into the estimation with these techniques improves the results as compared to GLSE according to the MSEP criterion (except for HOMBLUP in the univariate case).

Comparisons were also made when prior information is not compatible which is obviously the case in practical applications. Improvement conditions over sample predictor GLSE when the prior information is not compatible were derived. Results are summarized in Table 2.

In this section these results are interpreted within the scope of the proposed algorithm as we did in Chapter 3. First some additional information is given about these predictors. They are then compared with respect to each other under correct prior information using the MSEP matrix criterion. Similar comparisons are not definitive with respect to their corresponding MSEP matrices when the prior information is not compatible. In this case their corresponding improvement conditions can be used to decide when each estimator is suitable for the improved estimation of deformation parameters.

Now let us consider HOMBLIP. Using (27) and matrix identities given by Theorem A.9, this predictor can be written as

$$\hat{x}_1 = [I - N^{-1}(\bar{N}^{-1} + \mu_0\mu_0')^{-1}]\hat{x}_s = \alpha \cdot \hat{x}_s \qquad (114)$$

where

$$\bar{N} := N(I + \Sigma_{xx}N)^{-1} \qquad (115)$$

$$\alpha := I - N^{-1}(\bar{N}^{-1} + \mu_0\mu_0')^{-1} \qquad (116)$$

and $\hat{x}_s$ is the sample predictor. For $\Sigma_{xx} \to 0$, (114) reduces to BLE. Therefore, this estimator is similar to BLE. Indeed, for the univariate case the expected value of $\hat{x}_1$ is

$$E(\hat{x}_1) = \alpha \cdot \mu \qquad (117)$$

where

$$\alpha = [1 - (1 + (n/\sigma_u^2) \sigma_x^2 + (n/\sigma_u^2) \mu^2)^{-1}] \hat{x}_s \qquad (118)$$

Since $0 < \alpha < 1$ for all $\mu := E(x)$, this predictor provides a shrinkage of the sample predictor $\hat{x}_s$ of GLSE type, and it underestimates the estimates of the expected values of the random parameters. Hence it is a biased predictor in the sense of $E(\hat{x}_1) \neq \mu$.

The usefulness of this predictor depends on the availability of prior information μ_0 about μ and the covariance matrix Σ_{xx} of the random parameter vector. It was shown that if this information is available and it is strictly correct then HOMBLIP improves the estimates with respect to the sample estimator (Corollary 5.5.1). Since $\mu_0 = \mu$ cannot be realized in practice, improvements over the sample estimator is possible if

$$\mu' (\bar{N}^{-1} + 2\mu_0\mu_0')^{-1} \mu \leqslant 1 \qquad (119)$$

as a result of Theorem 5.2.2, or if

$$\mu' \bar{N} \mu \leqslant 1 \qquad (120)$$

which is an upper bound for (119). Under the assumptions given in Corollary 5.5.2, namely $u \sim (0, \sigma_u^2 I)$ and $x \sim (\mu, (\sigma_u^2/k)I)$, where $k := \sigma_u^2/\sigma_x^2$, this expression reduces to

$$\sum_{i=1}^{m} \frac{t_i^2/\sigma_u^2}{k^{-1} + \lambda_i} \leqslant 1 \qquad (121)$$

where t_i are the elements of $t = C' \mu$, and λ_i are the eigenvalues of $(A'A)^{-1} = C \Lambda C'$. The following deductions can be made in order for (121) to hold

i) Signal-to-noise ratios t_i^2/σ_u^2 should be sufficiently small. <u>Note that t_i is a function of $\mu := E(x)$.</u>

ii)   Components of signal vector should not be very pronounced when λ_i is small.

iii)   k should be sufficiently small (i.e., σ_X^2 is large).

Consider now HOMBLUP given by (59) which can also be expressed after a simple manipulation as

$$\hat{x}_2 = \left[I - N^{-1}\bar{N}^{\frac{1}{2}} \left(I - \frac{\bar{N}^{\frac{1}{2}}\mu\mu'\bar{N}^{\frac{1}{2}}}{\mu'\bar{N}\mu} \right) \bar{N}^{\frac{1}{2}} \right] \hat{x}_s \qquad (122)$$

Since the elements of the matrix within parentheses are bounded, HOMBLUP displays robust properties with respect to an incompatible prior information (i.e., when μ_0 which is used in (122) to replace unknown μ is not proportional to μ). This predictor is also unbiased in the sense of $E(\hat{x}_2) = \mu$. However, it reduces to the sample predictor $\hat{x}_s$ for the univariate application, thereby offering no improvements in this case.

If HOMBLIP is compared to HOMBLUP when $\mu_0 = \mu$ under Definition 2,

$$MSEP(\hat{x}_2) - MSEP(\hat{x}_1) = (\mu_0' \bar{N}\mu_0)^{-1} (1 + \mu_0' \bar{N}\mu_0)^{-1} \mu_0\mu_0' \geqslant 0 \qquad (123)$$

holds. HOMBLUP is therefore not as efficient as HOMBLIP over $\hat{x}_s$ under strictly correct prior information. On the other hand, if prior information is not compatible, then, as in the case of estimation with deterministic parameters, the superiority of one predictor over the other cannot be established directly using Definition 2 since the resulting difference matrix is indefinite again. In this case the improvement conditions given by

$$\mu'\bar{N}\mu \leqslant 1 \qquad (124)$$

$$\mu'(\bar{N}^{-1} + 2\mu_0\mu_0')^{-1}\mu \leqslant 1, \quad \mu_0 \neq \mu \qquad (125)$$

for HOMBLUP and HOMBLIP respectively are more informative. Since (123) is an upper bound for (124), the improvement condition for HOMBLIP is comparatively easier to achieve when $\mu_0 \gg 0$. Since (124) was used in

making inferences about (125) as an upper bound, the same results given by (121) hold also for HOMBLUP.

The third predictor, INHOMBLUP, which was examined in section 5.7 possesses different properties with respect to HOMBLIP and HOMBLUP. Using (99) and matrix identities given by Theorem A.9, this predictor can be written as

$$\hat{x}_3 = \hat{x}_s - N^{-1}\bar{N}(\hat{x}_s - \mu_0) \tag{126}$$

Compared to (114) and (122) this predictor is sensitive to prior information due to the unbounded effect of μ_0 in (126) (actually this statement is not strictly correct since the effect of μ_0 can be controlled artificially through Σ_{xx} as we shall examine later under certain simplifying assumptions). This is not desirable if the prior information is likely to be wrong. However large discrepancies between μ_0 and μ can easily be detected through the testing procedure given in section 5.2 according to the proposed approach, and prior information may either be corrected or not used at all.

In the case of small discrepancies this predictor can still be used to improve the estimation of stochastic parameters provided that improvement condition (104) derived in section 5.7 holds.

Before we elaborate on this condition it should be noted that this predictor, in addition to being unbiased, is better than the other two predictors with respect to Definition 2 when prior information is compatible. That is, from (31) and (100)

$$MSEP(\hat{x}_1) - MSEP(\hat{x}_3) = (1 + \mu'\bar{N}\mu)^{-1}(I + \Sigma_{xx}N)^{-1}\mu\mu'[(I + \Sigma_{xx}N)^{-1}]' \geqslant 0 \tag{127}$$

Hence it is better than HOMBLIP and also HOMBLUP as a result of (123).

Now consider the improvement condition (85) for INHOMBLUP when $\mu_0 \neq \mu$

$$s'\bar{N}s \leqslant 1 \tag{128a}$$

Under the simplifying assumptions $u \sim (0, \sigma_u^2 I)$ and $x \sim (\mu, (\sigma_u^2/k)I)$, this

condition reduces, as given in (112), to

$$s'[k^{-1}I + (A'A)^{-1}]^{-1}s < 1 \qquad (128b)$$

where $k: = \dfrac{\sigma_u^2}{\sigma_x^2}$ and σ_u^2, σ_x^2 are the a priori variances of the disturbances and the stochastic parameters respectively. Using Theorem A.5, $(A'A)^{-1} = C\Lambda C'$, (128b) can be written as

$$\sum_{i=1}^{m} \frac{t_i^2/\sigma_u^2}{k^{-1} + \lambda_i} < 1 \qquad (129)$$

where t_i are the elements of $t = C's$ (note that the $m \times 1$ t vector is a function of $m \times 1$ vector s of systematic deviations rather than the function of $\mu := E(x)$ as in the case of HOMBLIP and HOMBLUP). λ_i are the eigenvalues of $(A'A)^{-1}$. Now (129) is likely to hold when

 i) The bias-to-noise ratio $(s's/\sigma_u^2)$ is sufficiently small
 ii) λ_i are not small
iii) s_i is not too erroneous in directions where there is little infor mation (i.e., when λ_i is small s_i is also small)
 iv) k^{-1} is large (i.e., σ_x^2 is large)

Now let us compare these three predictors with respect to each other using their corresponding improvement conditions over GLSE when $\mu_0 \neq \mu$. If the following inequality is considered

$$\frac{1}{\lambda_{min}} \frac{t't}{\sigma_u^2} > \sum_{i=1}^{m} \frac{t_i^2/\sigma_u^2}{k^{-1} + \lambda_i} \qquad (130)$$

the improvement conditions for HOMBLIP and HOMBLUP given by (120) and for INHOMBLUP given by (129) can further be simplified to

$$\mu'\mu < (k^{-1} + \lambda_{min}) \sigma_u^2 \qquad (131a)$$

$$s's < (k^{-1} + \lambda_{min}) \sigma_u^2 \qquad (131b)$$

where λ_{min} is the minimum eigenvalue of $(A'A)^{-1}$. Since the left-hand sides of the above equations are the same, the preference of one estimator over the other is determined by the magnitude of the expected value of the random parameter vector μ and the magnitude of the systematic errors s.

Before we elaborate on this topic let us examine the other variables which may contribute to improvements over the sample predictor. Since in practice crustal movement observations such as repeated baseline measurements are performed using the most precise existing measurement techniques, σ_u^2 (the a priori variances of the observed baseline differences) can be considered to be fixed and obviously cannot be used as a control variable in achieving the improvement conditions. Similarly, λ_{min} (minimum eigenvalue of $(A'A)^{-1}$) would not play an important role in achieving the improvements since it is expected to be small as a result of the optimal design of deformation networks. For instance, the D-optimal design given in Chapter 3 renders the eigenvalues of $(A'A)^{-1}$ small, whereas the third variable k := σ_u^2/σ_x^2 can be used effectively to achieve the improvement conditions.

More uncertainty can be attributed to the prior information through σ_x^2. However, the larger σ_x^2 is the less the effect of prior information on the estimates will be. This in turn implies that the gain by using prior information would decrease. Although the gradual reduction of the effect of additional information on the estimates by this way is known in geodetic practice in general, the above conditions can be used effectively in reducing the effect of prior information efficiently.

Since the right-hand sides of the simplified improvement conditions (131a) and (131b) are the same, the usefulness of the predictors depends on the magnitudes of s and μ. Although these quantities are unknown in practice, the above improvement condition has already provided enough information about the central aspects of these predictors without resorting to extensive simulation studies in order to decide whether to incorporate prior information into the estimation of random deformation parameters in the improved estimation stage of the proposed algorithmic approach.

103

In the meantime, sample estimates $\hat{\mu}_s$ of μ computed, for instance, using the estimator given by Lemma 1 can be used to get some information about μ and s to evaluate the improvement conditions, provided that the null hypothesis testing is not rejected. In this case, if $\hat{\mu}_s'\hat{\mu}_s < \hat{s}'\hat{s}$, where ($\hat{s} := \hat{\mu}_s - \mu_0$), then HOMBLIP and HOMBLUP are preferable; otherwise INHOMBLUP is advantageous in introducing prior information. If null hypothesis testing is rejected, prior information should be introduced cautiously by attributing more uncertainty to priors or it should not be used at all.

In Chapter 4 the MSE matrix criterion for the estimation of deterministic parameters, and in this chapter its predictive version (MSEP), has been used to derive the improvement conditions with respect to the sample estimator GLSE as a result of noncompatible prior information. However, many different methods in comparing estimators are in use in statistics.

Among these, the unbiasedness property has not been given much attention in this study because even the unbiased estimators turn out to be biased when prior information is not compatible, as readily demonstrated in Chapters 4 and 5. Moreover, the unbiasedness property is not a measure of closeness to the true value especially when the number of observations is not very large.

Alternatively, the minimum variance criterion has not bee used since the variances of these estimators can always be made arbitrarily small by increasing artificially the amount of bias. Although the MSE and MSEP matrix criteria compensate for these deficiencies, there is no special reason to think that this is the best way of comparing estimators. It seems that they give too much weight to the case when the observations are not very precise or the prior information is badly wrong.

The Bayesian interpretation of these estimators is also possible (see, for example, (Mood et al., 1974) for a Bayesian derivation of INHOMBLUP for normally distributed random parameters).

Kubik (1986, private communication) has indicated equivariance properties of estimators in connection with changes in origin and scale

104

of measurements. These properties may have important ramifications in the selection of estimators depending on the type of observations and the nature of the estimates. They are defined as follows.

If y_1, ..., y_n represent measurements and a parameter being estimated are in the same units with these measurements, it is reasonable to require that an estimator $\hat{x}$ satisfies the following property,

$$\hat{x}(y_1 + c, \ldots, y_n + c) = \hat{x}(y_1, \ldots, y_n) + c \tag{132}$$

for every value of y_1, y_2, ..., y_n and constant c. In other words, the estimate should increase by an amount c if each measurement increases by c. This property is also known as location invariance.

Similarly an estimator $\hat{x}$ is said to be scale equivariant (scale invariant) if

$$\hat{x}(cy_1, \ldots, cy_n) = c\hat{x}(y_1, \ldots, y_n) \tag{133}$$

for every y_1, ..., y_n and constant c. The idea is that the estimator should be independent of measurement units employed.

Now let us consider the following linear model

$$y_i = x + u_i \qquad i = 1, \ldots, n$$
$$u_i \sim (0, \sigma_u^2) \tag{134}$$

for a location estimator of x where y_i denotes observations, u_i denotes identically and independently distributed disturbances with a priori variance σ_u^2 and x is the parameter to be estimated (stochastic or deterministic). Then the sample mean $\overline{x}$, for instance, is a scale invariant estimator since

$$c\overline{x} = \sum_{i=1}^{n} \frac{cy_i}{n} = c\frac{\sum_{i=1}^{n} y_i}{n} = c\overline{x} \tag{135}$$

105

Similarly if a priori information about x is available, BLE of x under the above linear model is given by

$$\hat{x}_1 = (1 + \frac{\sigma_u^2}{x_0^2 n})^{-1} \bar{x} \tag{136}$$

This estimator is also scale equivariant (in this case changes in scale affect x_0 and σ_u^2 also) because

$$c\hat{x}_1 = (1 + \frac{c^2 \sigma_u^2}{c^2 x_0^2 n})^{-1} c\bar{x} = c\bar{x} \tag{137}$$

The scale equivariance of the other estimators and predictors can be demonstrated in a similar way.

The sample mean $\bar{x}$ is also a location equivariant estimator since, using (132),

$$\hat{x} + c = \frac{\sum_{i=1}^{n} (y_i + c)}{n} = \frac{\sum_{i=1}^{n} y_i}{n} + c = \hat{x} + c \tag{138}$$

BLUUE, BLUE, HOMBLUP and INHOMBLUP are also location equivariant under the linear model (134). However, BLE and HOMBLIP are not location equivariant. The following proposition, which can be easily proven using the principle of reductio ad absurdum, generalizes the demonstration to different estimators: "If $\hat{x}$ is a location equivariant estimator of x then no multiple of $\hat{x}$ satisfies location equivariance property." Therefore BLE and HOMBLIP cannot be location equivariant because they can be expressed as shrinkage of the sample mean which is location equivariant.

In terms of the improvement conditions for BLE and HOMBLIP this implies that these conditions can always be satisfied or dissatisfied by the proper choice of origin of measurements. However in crustal movement applications where measurements are generally differences of baseline observations or angles, any changes in the origin of these observations do not make much sense. Hence we need not insist on the location equivariance in these cases. In those applications where the

origin change is meaningful, these two estimators should be carefully handled because the estimates can arbitrarily be changed when the origin of the measurements is changed.

Chapter 6

CONCLUSIONS AND RECOMMENDATIONS

In recent years, the analysis of crustal deformation measurements has become important as a result of current improvements in the geodetic methods and an increasing amount of theoretical and observational data provided by several earth sciences. However, a combined analysis of different types of information has not received proper attention.

In this study, a "first generation" data analysis algorithm which combines extraneous information with current geodetic measurements was proposed. Relevant methods which can be used in the algorithm have been discussed. Prior information is the unifying feature of this algorithm. Some of the problems which may arise through the use of additional information in the analysis have been indicated and preventive measures were demonstrated.

The first step of the algorithm is the optimal design of the geodetic networks. Although the current research on optimal designs of geodetic networks is extensive, deformation model oriented network designs have yet to be exploited. The duality between geodetic axiomatic model oriented designs and deformation model oriented designs deserves further investigations. As an example to deformation model oriented designs, it was shown that the regular polygonal deformation networks composed of equilateral triangles are uniformly D-optimal for homogeneous deformation field. The methodology used in this derivation can easily be extended to the optimal design of possible alternative deformation models which may arise in practice.

The concept of optimal network designs, which does well for the totality of different postulated models, was identified. Such designs can in general be constructed iteratively using prior preferences as weights

108

on each model. Formulation of the problem for deformation networks is open to further investigations.

The second step in the algorithm identifies the descriptive model of the deformation field. Previous information about the nature of crustal movements as a result of previous measurements and/or theoretical considerations may suggest more than one model to represent the deformations. In this case, sequential experimental designs are necessary to identify the suitable model. A method based on the entropy measure of information, proposed by Box and Hill (1967), was applied to a group of postulated deformation models and the identification of the correct model was demonstrated through an example. The method uses prior information in the form of prior preferences for each model in a Bayesian setting. This additional information, in turn, allows prediction of optimal observations and computation of the likelihood of each model. Although the numerical example indicated that the method effectively identifies the correct model, its usefulness needs to be demonstrated in practice.

The next step in the algorithm is the improved estimation of deformation parameters. Although these parameters are estimated in the process of model discrimination, they can further be improved by the use of extraneous information about them. Compared to the previous topics and the regular estimation techniques, the use of additional information in the estimation of deformation parameters has not been exploited in detail in crustal movement analysis. Therefore, a major part of this study has been devoted to this subject.

According to the proposed algorithm, prior information must first be checked through null-hypothesis testing against the estimates calculated using the least squares method, which employs only sample observations, before it is introduced to the final estimation. This procedure is likely to detect large discrepancies between two different estimates, but may not be conclusive if both estimates are generally in the same direction and close in magnitude. However, it was demonstrated that even in these cases, incompatible prior information can still be used to improve the final estimates under certain circumstances.

Introduction of prior information can be achieved using different estimation techniques. Schaffrin (1983) proposed a group of estimators that can be used for such purposes. Of these techniques, Best Linear Uniformly Unbiased Estimator BLUUE, Best Linear Unbiased Estimator BLUE, and Best Linear Estimator BLE, were examined due to their rather unknown statistical properties in the geodetic area.

Since the purpose of using prior information is to improve the estimates in the proposed algorithm, these estimators were analytically compared against GLSE. It was shown that they are in general better than GLSE with respect to their corresponding mean square error matrices. However, these inferences can only be made as long as prior information is known adequately. Otherwise, possible discrepancies between prior information and the true value of parameters render BLUUE and BLUE biased and introduces additional biases in the biased estimator BLE. Under these circumstances the advantage of using additional information was investigated. Comparisons of these estimators under spurious prior information against GLSE led to the conditions for improvements (Table 1). These results also provide the means to select a suitable estimator under wrong prior information. In general, improvements are possible for biased BLUUE if prior information is not too erroneous or observations relatively poor. Meanwhile, it was shown that BLUE and BLE are robust against wrong prior information. If the true values of the parameters are relatively small with respect to the observation noise, then improvements over GLSE are possible by using these estimators.

As discussed in Chapter 5, deformation parameters can be regarded as random quantities with certain means and variances. This interpretation led to the study of estimation of stochastic parameters (prediction). In this class, Best Inhomogeneously Linear Unbiased Prediction INHOMBLUP, Best Homogeneously Linear Unbiased Prediction HOMBLUP and Best Homogeneously Linear Prediction HOMBLIP were considered. A null hypothesis testing procedure was developed to check the compatibility of prior information about the means of the stochastic deformation parameters with the estimates implied by the sample

observations. It was demonstrated that, as a reference criterion, GLSE is also a Best Homogeneously Linear Unbiased Prediction if no prior information is available. The above predictors were then compared against GLSE using the mean square matrix of prediction criterion assuming that correct prior information on the mean of the true stochastic parameter vector and its covariance matrix are available. In each case, it was shown that prior information improves the estimation of stochastic parameters.

Following the similar discussions in the case of estimation of deterministic parameters with erroneous prior information, once again the conditions for improvements were derived (Table 2). It was found that these predictors not only possess the properties of the estimators with deterministic parameters examined, but they also enjoy the additional flexibility provided by the stochastic interpretation of deformation parameters through their covariance matrices.

Finally, analytic conditions derived for these estimators and predictors are not only useful for understanding their properties but are also important in deciding which estimator or predictor is to be used in the improved estimation stage of the proposed algorithm. These conditions also replace extensive simulation studies for different applications. Particular cases can easily be inferred from the general conditions (Tables 1 and 2).

Appendix
SOME USEFUL THEOREMS AND IDENTITIES

__Theorem A.1:__ Let I be the n×n identity matrix and g an n×1 vector. Then

$$I - gg' \geq 0 \qquad iff \qquad g'g \leq 1$$

For proof see (Yancey et al., 1974).

__Theorem A.2:__ Let A be an n×n symmetric matrix; if $A \geq 0$, then A can be written as

$$A = A^{\frac{1}{2}}A^{\frac{1}{2}}$$

and $A^{\frac{1}{2}}$ is also symmetric. For proof see (Mardia et al., 1979).

__Theorem A.3:__ Let A be an n×n matrix and U an n×1 vector. Then

$$\sup_x \frac{(U'x)^2}{x'Ax} = U'A^{-1}U$$

and the supremum is attained at $x = A^{-1}U$. For proof see (Rao, 1973).

__Theorem A.4:__ Let A be an n×n matrix and $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_n$ be its eigenvalues then

$$\sup_x \frac{x'Ax}{x'x} = \lambda_1 \qquad , \qquad \inf_x \frac{x'Ax}{x'x} = \lambda_n$$

For proof see (Rao, 1973).

__Theorem A.5:__ Let A be an n×n symmetric matrix. Then there exists an orthogonal matrix P such that

$$P'AP = \Lambda$$

with $P'P = PP' = I$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ are the eigenvalues of A. For proof see (Mardia et al., 1979).

Theorem A.6: Let A be an n×n p.s.d. matrix; then PAP' is p.s.d. for any n×n matrix P. For proof see (Graybill, 1976).

Theorem A.7: If $A > 0$ and $B \geq 0$, then $A^{-1} - (A + B)^{-1}$ is p.s.d. For proof see (Goldberger, 1964).

Theorem A.8: Let A be an n×n symmetric matrix; then

$$\mathrm{tr}\{A\} = \sum \lambda_i$$

where λ_i are the eigenvalues of A. For proof apply theorem A.5.

Theorem A.9: If A is an n×n matrix, p.d. and $(A + BDC) > 0$ and BDC are conforming matrices then

$$(A + BDC)^{-1} = A^{-1} - A^{-1}BDCA^{-1}(I + BDCA^{-1})^{-1}$$

$$= A^{-1} - (I + A^{-1}BDC)^{-1}A^{-1}BDCA^{-1}$$

holds. For the derivation see (Henderson and Searle, 1981). Useful extensions of this theorem in the general notation of the text are

$$A'(\Sigma_{uu} + A\Sigma_{xx}A')^{-1}A = N - N(I + \Sigma_{xx}N)^{-1}\Sigma_{xx}N =$$

$$= N(I + \Sigma_{xx}N)^{-1} = (I + N\Sigma_{xx})^{-1}N =: \bar{N}$$

and for a conforming vector x

$$(N^{-1} + xx')^{-1} = N - \frac{N\,xx'N}{1 + x'Nx}$$

where

$$N := A'\Sigma_{uu}A$$

Theorem A.10: If $x \sim N_p(\mu, \Sigma)$, $\Sigma > 0$, then

113

$$x'\Sigma^{-1}x \sim \chi_p^2(\mu'\Sigma^{-1}\mu)$$

$$(x - \mu)'\Sigma^{-1}(x - \mu) \sim \chi_p^2$$

For proof see (Tautenburg, 1982).

Theorem A.11: Let p×1 random vector $x \sim N(\mu, \Sigma)$ where Σ is full rank. Let A be a q×p matrix of constants, and let b be any q×1 vector of constants. Then

$$y = Ax + b \sim N_p(A\mu + b, \ A\Sigma A')$$

For proof, see (Graybill, 1976).

REFERENCES

Baarda, W., 1977: "Reliability and Precision of Networks," paper to VII Int. Course for Eng. Surveys of High Precision, Darmstadt, FRG.

Baarda, W., 1978: "Difficulties in Establishing a Model for Testing Crustal Movements," Progress in Geodynamics, Proc. National Symp. on Geodynamics, Amsterdam, April 1975, North Holland Pub. Amsterdam/New York.

Beloussov, V.V., 1979: "Why do I not accept plate tectonics?" EOS, Transactions AGU, Vol. 60, pp. 207-211.

Bibby, J. and H. Tautenburg, 1977: Prediction and Improved Estimation in Linear Models, John Wiley and Sons, New York.

Bock, Y., 1982: "The Use of Baseline Measurements and Geophysical Models for the Estimation of Crustal Deformations and the Terrestrial System," Ohio State Univ. Dept. of Geodetic Science and Surveying Rep. 337, Columbus.

Box, G.E.P. and W.J. Hill, 1967: "Discrimination Among Mechanistic Models," Technometrics, Vol. 9, No. 1, pp. 57-71.

Brunner, F.K., 1979: "On the Analysis of Geodetic Networks for the Determination of Incremental Strain Tensor," Survey Review, Vol. 192.

Brunner, F.K., R. Coleman and B. Hirsch, 1980: "A Comparison of Computation Methods for Crustal Strains from Geodetic Measurements," Tectonophysics, Vol. 71, pp. 281-2, 298.

Chrzanowski, A., 1981: "A Comparison of Different Approaches into the Analysis of Deformation Measurements," FIG, XVI, International Congress, Switzerland.

Chrzanowski, A., Y. Q. Chen Qi and J. M. Secord, 1983: "A Generalized Approach to the Geometrical Analysis of Deformation Surveys," Proc. of the Third Int. Symp. on Deformation Measurements by Geodetic Methods, I. Joö and A. Detrekoi (Eds.).

Coates, J.R., H. Frey, J. Bosworth and G.D. Mead, 1985: "Space Age Geodesy: The NASA Crustal Dynamics Project," NASA Crustal Dynamics Project Preprint No. 51, GSFC, MD.

115

Cook, D.R. and Nachtsheim C., 1982: "Model Robust Linear Optimal Designs," Technometrics, Vol. 24.

Davidson, J. and O. Trask, 1981: "Utilization of Mobile VLBI for Geodetic Measurements," IEEE Trans. G.R.S.

Drake, C.L., 1983: "International Geoscience Past and Future," EOS Transactions AGU, pp. 961,963.

Dermanis, A. 1981: "Geodetic Estimablity of Crustal Deformation Parameters," Quatern. Geod., Vol. 2, pp. 159-169.

Drewes, H., 1982: "A Geodetic Approach for the Recovery of Global Kinematic Plate Parameters," Bull. Géod., Vol. 56, No. 1, pp. 72-79.

Ellmer, W. and W. M. Welsh, 1982: "Some Consideration on the Interpretation in Series of Geodetic Measurements for Monitoring Crustal Movement Process," 4th Canadian Symp. on Mining Surveying and Deformation Measurements, Alberta.

Farebrother, R.W., 1976: "Further Results on the Mean Square Error of Ridge Regresion," J.R. Statis. Soc., B38, pp. 248-250.

Fedorov, V.V., 1972: Theory of Optimal Experiments, Academic Press, New York.

Goldberger, A.S., 1962: "Best Linear Unbiased Prediction in the Generalized Regression Model," JASA, Vol. 57, pp. 369-375.

Goldberger, A.S., 1964: Econometric Theory, John Wiley and Sons Inc., New York.

Goldstein, H., 1965: Classical Mechanics, Addison-Wesley Publ.

Grafarend, E.W., 1977: "Operational Geodesy," Approximation Methods in Geodesy, Lecture Delivered at the Second International Summer School in Physical Geodesy, Ramsau.

Grafarend, E.W., H. Heister, R. Kelm, H. Kropff and B. Schaffrin, 1979: Optimierung Geodatischer Messoperationen, H. Wichman Verlag, Karlsruhe, FRG.

Graybill, F.A., 1976: Theory and Application of Linear Model, Duxbury Press, North Scituate, MA.

Harvey, B.R, 1985: "The Combination of VLBI and Ground Data for Geodesy and Geophysics," School of Surveying, The University of South Wales Rep. S27.

Harville, D., 1976: "Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects," The Annals of Statistics, Vol. 4, No. 2, pp. 384-395.

Heck, B., 1980: "Statische Ausreisserkriterien zur Kontrolle Geodatischer Beobachtungen," VII Int. Kurs fur Ingenieurvemessung, Zurich B 10.

Henderson, H.V. and S.R. Searle, 1981: "On Deriving the Inverse of a Sum of Matrices," SIAM Rev., Vol. 23, pp. 53-60.

James, W. and C. Stein, 1961: "Estimation with Quadratic Loss," Proc. Fourth Berkeley Symp. in Mathematical Statistics and Probability, 1, pp. 361-379.

Kaula, W.M., 1975: "Absolute Plate Motions by Boundary Velocity Minimization" J. Geophys. Res., Vol. 80, No.2, pp. 244-248.

Kaula, W.M., 1978: "Geodynamics Problems," in application of Geodesy to Geodynamics, Proc. of 9th GEOP Conference, I.I. Mueller (Ed.), Ohio State Univ. Dept. of Geodetic Science and Surveying Rep. 280, pp. 345-351, Columbus.

Kiefer, J., 1961: "Optimum Designs in Regression Problems II," Ann. Math. Statist., Vol. 32, No. 298.

Koch, K.R. and D. Fritsch, 1980: "Multivariate Hypothesis Tests for Detecting Recent Crustal Movements," Tectonophysics, Vol. 71, pp. 301-313.

Kok, J.J., 1977: "The B-Method of Testing Applied to RETRIG Computations," Rep. of the Symp. of IAG Subcom. for RETRIG, Brussels, pp. 106-116.

Kok, J.J., 1980: "The 1979 Adjustment of the UELN and Its Analysis of Reliability and Precision," Proc. NAD 1980 Symp., Ottawa, pp. 455-483.

Kullback, S., 1959: Information Theory and Statistics, John Wiley and Sons, Inc., New York.

Läuter H., 1974: "Experimental Planning in a Class of Models," Mathematische Operationfoschung und Statist., Vol. 5, pp. 379-397.

Livieratos, E., 1980: "Crustal Deformations from Geodetic Measurements," Boll. Geofis. Teor. Applic., Vol. 22, pp. 255-260.

Lyzenza G.A., K.S. Wallace, J.L. Fanselov, A. Raefsky, P.M Groth, 1985: "Tectonic Motions in California Inferred from VLBI Observations, 1980-1984," JPL Geodesy and Geophysics Preprint No. 129, Pasedena, California.

Mardia, K.V., J.T. Kent, J.M. Bibby, 1979: <u>Multivariate Analysis</u>, Academic Press, London.

Mierlo, V.J., 1975: "Statistical Analysis of Geodetic Networks Designed for Detection of Crustal Movements," Progress in Geodynamics, Proc. National Symp. on Geodynamics, Amsterdam, April 1975, North Holland Publ., New York/Amsterdam.

Mierlo, V.J., 1978: "A Testing Procedure for Analysing Geodetic Deformations," Proc. of the Second Int. Symp. on Deformation Measurements by Geodetic Methods, Bonn.

Minster J.B. and T.H. Jordan, 1978: "Present Day Plate Motions," J. Geophys. Res., Vol. 83, No. 11, pp. 5331-5354.

Molenaar, M., 1981: "A Further Inquiry into the Theory of S-Transformations and Criterion Matrices," Netherlands Geodetic Commission, Publ. on Geodesy, Vol. 7, No. 10.

Mood, M.A., F.A. Graybill and D.C. Boes, 1974: <u>Introduction to the Theory of Statistics</u>, 3rd ed., McGraw-Hill, New York.

National Aeronautics and Space Administration (NASA), 1978: "Application of Space Technology to Crustal Dynamics and Earthquake Research," NASA Technical Paper 1464, Office of Space and Terrestrial Applications, Washington, D.C.

National Research Council - Committee on Geodesy, 1981: "Geodetic Monitoring of Tectonic Deformation - Toward a Strategy," National Academy Press, Washington, D.C.

Ney, B., 1978: "Determination of Deformations of Horizontal Geodetic Networks by Approximation," Proc. Int. Symp. on Optimization of Designs and Computation of Control Networks, Sopron, 1977.

Niemeier, W., 1979: "Zur Congruenz Mehrfach Beobachteter Geodatischer Netze," Wiss. Arb. der Fachrichtung Vermessungwesen der Universitat, Hannover, Nr. 88.

Papoulis, A., 1984: <u>Probability, Random Variables and Stochastic Processes</u>, McGraw Hill, New York.

Pelzer, H., 1974: "Nevere Ergebnisse bei der Statistichen Analyse von Deformationsmessungen," XIV FIG Congress, Washington, paper 608.3.

Pope A.J., 1966: "Strain Analysis of Repeated Triangulation for the Investigation of Crustal Movement," Masters Thesis, Dept. of Geodetic Science and Surveying, Columbus.

Popper, K., 1972: <u>Objective Knowledge</u>, Oxford University Press, London.

Rao, C.R., 1965: "The Theory of Least Squares when the Parameters are Stochastic and its Application to the Analysis of Growth Curves," Biometrica, 52, 3-4, pp. 447-458.

Rao, C.R., 1973: Linear Statistical Inference and its Applications, John Wiley and Sons, Inc., N.Y.

Schaffrin, B., 1983: "Model Choice and Adjustment Techniques in the Presence of Prior Information," Ohio State Univ. Dept. of Geodetic Science Rep. 351, Columbus.

Searle, S.R., 1974: "Prediction, Mixed Models and Variance Components," Reliability and Biometry, F. Proschan, R.J. Serfling (Eds.), SIAM, Philadelphia, pp. 229-266.

Searle, S.R., 1982: Matrix Algebra, Useful for Statistics, John Wiley and Sons, New York.

Shannon, C.E., 1948: "A Mathematical Theory of Communication," The Bell Syst. Tech. J., Vol. XXVII, No. 3, pp. 379-657.

Silvey, S.D., 1980: Optimal Design, Chapman and Hall, New York/London.

Slater, L.E., A. McGarr, J.O. Laugbein and M.F. Linker, 1983: "Multiwavelength EDM Measurements in Southern California," Tectonophysics, Vol. 97, pp. 39.

Solomon, S.C. and N.H. Sleep, 1974: "Some Simple Physical Models for Absolute Plate Motions," J. Geophys. Res., Vol. 79, No. 17, pp. 2557-2567.

Tautenburg, H., 1982: Prior Information in Linear Models, John Wiley and Sons, New York.

Terasvirta, T., 1979: "Some Results on Improving the Least Square Estimation of Linear Models by Mixed Estimation," Discussion Paper 7914, Louvain, CORE.

Thatcher, W., 1978: "Crustal Movement and Earthquake Related Deformations," U.S. Geological Survey, California.

Thatcher, W., 1983: "Non-linear Strain Build-up and Earthquake Cycle on the San Andreas Fault," J. Geophys. Res., Vol. 88, pp. 5893-5902.

Theil, H. and Goldberger, A.S., 1961: "On Pure and Mixed Estimation in Economics," Int. Economics Rev., Vol. 2, pp. 65-78.

Theil, H., 1963: "On the Use of Incomplete Prior Information in Regression Analysis," Amer. Stat. Assoc. J., pp.407-414.

Ullah, A. and S. Ullah, 1978: "Double k-Class Estimators of Coefficients in Linear Regression," Econometrica, 46.

Uotila, U.A., 1980: "Advanced Adjustment Computations," Lecture Notes, Ohio State Univ. Dept. of Geodetic Science and Surveying, Columbus.

Vinod, H.D., 1976: "Simulation and Extension of a Minimum MSE Estimator in Comparison with Stein's," Technometrics, pp. 491-496.

Voievodine, V., 1976: "Algèbre Linéaire," Editions MIR, Moscou.

Welsch W., 1982: "Description of Homogeneous Strains and Some Remarks to their Analysis," Proc. IAG Symp. on Geodetic Networks and Comp., R. Sigl (Ed.), Publ. DGK B-58/v, Munchen, 1982, pp. 188-205.

Welsch W., 1983: "Finite Element Analysis of Strain Patterns from Geodetic Observations Across a Plate Margin," Tectonophysics, Vol. 97, pp. 57-71.

Yancey, T.A., Judge, G.G. and Bock, M.E., 1974: "A Mean Square Error Test when Stochastic Restrictions are Used in Regression", Commun. Statist., Vol. 3, pp. 755-768.